৩৯৩/৭.৯.৭০

SANKHYĀ : THE INDIAN JOURNAL OF STATISTICS

CONTENTS OF VOLUME TWENTY, 1958

(*All rights reserved*)

PAPERS

Page

On moments and cumulants of systems of statistics. *By G. S. James* ... 1

A comparison of serological and somatometrical methods used in differentiating between certain East African Racial Groups, with special reference to D^2 Analysis. *By D. A. East and L. Oschinsky* ... 31

Science and National Planning. *By P. C. Mahalanobis* 69

The scope of biological statistics. *By J. B. S. Haldane* 195

Examples of inconsistency of maximum likelihood estimates. *By R. R. Bahadur* 207

Maximum likelihood estimation for the multinomial distribution with infinite number of cells. *By C. Radhakrishna Rao* 211

A remark on strong measurability. *By V. S. Varadarajan* 219

A useful covergence theorem. *By V. S. Varadarajan* 221

On statistics independent of sufficient statistics. *By D. Basu* 223

On configurations and non-isomorphism of some incomplete block designs. *By M. Atiqullah* 227

Tables of random normal deviates. *By J. M. Sengupta and Nikhilesh Bhattacharya* 249

On sampling with and without replacement. *By D. Basu* 287

Diffusion by discrete movements. *By H. C. Gupta* 295

On the distribution of the maximum value of an equally correlated sample from a normal population. *By A. Kudô* 309

Hypergeometric expansions for incomplete moments of the bivariate normal distribution. *By A. R. Kamat* 317

Incomplete moments of the trivariate normal distribution. *By A. R. Kamat* ... 321

A note about a family of Perks' distribution. *By Joseph Talacko* 323

Recent experiments in statistical sampling in the Indian Statistical Institute. *By P. C. Mahalanobis* 329

REPORT

Indian Statistical Institute : Twenty-fifth Annual Report 109

BUREAU OF EDNL. & PSYL.
LIBRARY

CORRIGENDA

Some observations on Input-output Analysis. *By Oscar Lange* 107
A note on Distance between Two Populations. *By S. R. Adke* 108

AUTHOR INDEX

Adke, S. R. Corrigenda to a paper on "A note on Distance between Two Populations" 108

Atiqullah, M. On configurations and non-isomorphism of some incomplete block designs 227

Bahadur, R. R. Examples of inconsistency of maximum likelihood estimates ... 207

Basu, D. On statistics independent of sufficient statistics 223

——— On sampling with and without replacement 287

Bhattacharya, Nikhilesh and Sengupta, J. M. Tables of random normal deviates 249

East, D. A. and Oschinsky, L. A comparison of serological and somatometrical methods used in differentiating between certain East African Racial Groups, with special reference to D^2 Analysis 31

Gupta, H. C. Diffusion by discrete movements 295

Haldane, J. B. S. The scope of biological statistics 195

James, G. S. On moments and cumulants of systems of statistics 1

Kamat, A. R. Hypergeometric expansions for incomplete moments of the bivariate normal distribution 317

——— Incomplete moments of the trivariate normal distribution ... 321

Kudô, A. On the distribution of the maximum value of an equally correlated sample from a normal population 309

Lange, Oscar. Corrigenda to a paper on "Some observations on Input-output Analysis" 107

Mahalanobis, P. C. Science and National Planning 69

——— Recent experiments in statistical sampling in the Indian Statistical Institute 329

Oschinsky, L. and East, D. A. A comparison of serological and somatometrical methods used in differentiating between certain East African Racial Groups, with special reference to D^2 Analysis 31

Rao, C. R. Maximum likelihood estimation for the multinomial distribution with infinite number of cells 211

Sengupta, J. M. and Bhattacharya, Nikhilesh. Tables of random normal deviates 249

Talacko, Joseph. A note about a family of Perks' distribution 323

Varadarajan, V. S. A remark on strong measurability 219

——— A useful convergence theorem 221

Bureau Edn. Psy. Research
DAVID HARE TRAINING COLLEGE
Dated 7·9·70

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

Vol. 20, Parts 1 & 2 September 1958

ON MOMENTS AND CUMULANTS OF SYSTEMS OF STATISTICS

By G. S. JAMES

University of Leeds, England

SUMMARY. A direct proof of Fisher's sampling rules for k-statistics is presented, in such a form that their generalization for other systems of statistics becomes apparent.

1. Introduction

The work we are about to discuss had its origin in the problem of determining standard errors and higher moments of the sampling distributions of moment-statistics. The derivation of these results is straightforward but tedious and the answers are very complicated in all but the simplest cases. (See, for example, Tchouproff (1919, 1921), Church (1925)).

A great step forward was made by Fisher (1930). He introduced certain functions of the sample values which he called k-statistics, whose expectations are exactly equal to the cumulants of the (infinite) population. He found that if the sampling cumulants of the k-statistics were expressed in terms of the population cumulants the formulae were very much simplified; more important, he gave, in a series of rules, an elegant combinatorial procedure for arriving at these results. He also gave an outline of a proof of the rules. Sukhatme (1938) wrote a paper in order to "... trace the origin of his [Fisher's] method and to describe the parallel combinatorial method of evaluating the coefficient of the moment term in the moment of the distribution of powers and products of any k's." This he did by means of a particular numerical examples, which, while they greatly help the student to appreciate Fisher's thought-processes when framing the rules, were not intended, it

OF EDNL. & PSYL.
LIBRARY

is believed, to furnish a full proof of them. M. G. Kendall (1940) developed a novel operational approach and used it (Kendall, 1940A) to give a new proof. While sufficiently neat, the approach is largely descriptive, and much is left to the reader's imagination—much that at least one reader finds difficult.

One object of the present paper is to present a direct proof, which it is hoped is both reasonably rigorous and easier to follow than earlier ones. In spite of the fact that simple explanatory examples have been introduced in various places, it will probably be helpful in practice if the reader has first read Chapter 11 of Kendall's (1947) book in order to get a working knowledge of the rules (which as Kendall remarks, are not as difficult as they look).

The second object of the paper is to consider how the results are modified when we consider the sampling properties of a general set of statistics $z_1, z_2, \ldots,$ (being homogeneous polynomial symmetric functions of the sample values, of degrees 1, 2,..., respectively) instead of the set of k-statistics. This problem was partially considered by Kendall (1942). His purpose was to see whether a set of sample statistics could be found which would obey even simpler rules than Fisher's, and this conclusion was in the negative, provided that one also required the new statistics to be seminvariant (*i.e.* absolutely invariant under a change of origin). However, he did not actually formulate all the rules for general statistics, although he did come very near to doing so. There is nothing in this paper which throws any doubt on Kendall's conclusion about the central position occupied by the k-statistics. Why, then, should we investigate other statistics than the k's ?

I think one answer was given by Sukhatme (1944). He said, "Practical statisticians have been slow to adopt Fisherian results into regular use. Several instances may be quoted where statisticians have preferred to transform his results into direct moments for purposes of practical application... . This situation is probably partly explained by the nature of k-statistics themselves which are slightly more difficult to comprehend and lack the simplicity of expression of moment-statistics."

Another answer is that in theoretical investigations we sometimes need expressions for the cumulants of sets of statistics other than the k's ; thus the author recently required joint cumulants for the rather odd series of statistics, $z_r = \bar{x}^r$ $(r = 1, 2, \ldots)$. Despite the fact that $z_r = z_1^r$, the general theory will deal with this.

Sukhatme (1944) worked out, by pedestrian methods, formulae for the moments of the sample moments $(m_1' = \bar{x};\ m_r = \Sigma(x_i - \bar{x})^r/n,\ r = 2, 3, \ldots)$ in terms of population moments, for samples from either a finite or infinite population, recording most of them up to order eight. Other authors have considered the problem of expressing the cumulants of the sample moments in terms of population cumulants. For example, Craig (1928) used direct methods, St. Georgescu (1932) used "associated functions" and a combinatorial method rather different from Fisher's, while Wishart (1933) modified Fisher's results by an operational method.

The extension of the work to multivariate sampling has been considered right from the beginning (Fisher, 1930), two main approaches being discernible : that in which we start with results in as many variates as possible and then specialize to get univariate, bivariate, ... results by equating variates, and that in which we start from a univariate formula and generalize it by processes of symbolic operation. The first approach is represented by Fisher, and in more modern times by Kaplan (1952), who made it more tractable by using "tensor" notation. The second is represented by Kendall (1940B). There is also the direct approach, again represented by Kendall (1947), in which bivariate results, for example, are derived by considering partitions and bipartitions of bipartite numbers, instead of ordinary numbers. While a description of the latter appears in § 3 of this paper, our proofs are mostly based on the Fisher-Kaplan approach, univariate and less general multivariate results being obtained as special cases, because this process seems to simplify the proofs very considerably.

Most of the recent work on the general theory has been in attempting to generalize the results to the case of sampling from a finite population, but we shall not touch on this very interesting problem here.

2. THE PROBLEMS AND RULES

Let $x_1, \ldots, x_n$ be a sample of size n, drawn from an infinite population whose moments about the origin are $\mu'_r = Ex^r$, and whose moments about the mean are

$$\mu_r = E(x-\mu'_1)^r = \sum_{j=0}^{r-2} (-)^j \binom{r}{j} \mu'_{r-j}\mu'^{j}_1 + (-)^{r-1}(r-1)\mu'^{r}_1. \qquad \ldots \quad (2.1)$$

The cumulants, κ_r, of the population are defined by the formal relation

$$\sum_{r=1}^{\infty} \frac{\kappa_r \theta^r}{r!} = \log \sum_{r=0}^{\infty} \frac{\mu'_r \theta^r}{r!} \qquad \ldots \quad (2.2)$$

between the cumulant-generating and moment-generating functions, whence

$$\kappa_r = \sum_{(a)\mathrm{p}r} (-)^{\alpha-1}(\alpha-1)! \left[\begin{matrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{matrix}\right] \mu'^{\alpha_1}_{a_1} \ldots \mu'^{\alpha_A}_{a_A}; \qquad \ldots \quad (2.3)$$

where

$$\left[\begin{matrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{matrix}\right] = \frac{r!}{(a_1!)^{\alpha_1}\alpha_1! \ldots (a_A!)^{\alpha_A}\alpha_A!} \qquad \ldots \quad (2.4)$$

is a "reduced multinomial coefficient," and $\sum\limits_{(a)pr}$ denotes the sum over all the partitions $(a_1^{\alpha_1} \ldots a_A^{\alpha_A})$, or (a) for conciseness, of the number r, with $\Sigma\, \alpha_j a_j = r$, $\Sigma\, \alpha_j = \alpha$. (We shall always denote partitions by triplets of lower-case, Greek and capital letters such as a, α, A). Similarly

$$\mu'_r = \sum_{(a)pr} \begin{bmatrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{bmatrix} \kappa_{a_1}^{\alpha_1} \ldots \kappa_{a_A}^{\alpha_A} . \qquad \ldots \quad (2.5)$$

Fisher (1930) defined the statistic k_r as that polynomial symmetric function of the sample values $x_1, \ldots, x_n$ whose expectation is κ_r; hence

$$k_r = \sum_{(a)pr} (-)^{\alpha-1}(\alpha-1)! \begin{bmatrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{bmatrix} \frac{\Sigma'(x_{i_1} \ldots x_{i_{\alpha_1}})^{a_1} \ldots (x_{i_{\alpha-\alpha_A+1}} \ldots x_{i_\alpha})^{a_A}}{n^{(\alpha)}}$$

$$= \sum_{(a)pr} (-)^{\alpha-1}(\alpha-1)! \begin{pmatrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{pmatrix} \frac{\Sigma(x_1 \ldots x_{\alpha_1})^{a_1} \ldots (x_{\alpha-\alpha_A+1} \ldots x_\alpha)^{a_A}}{n^{(\alpha)}} \qquad \ldots \quad (2.6)$$

where

$$n^{(\alpha)} = n(n-1) \ldots (n-\alpha+1), \qquad \ldots \quad (2.7)$$

$$\begin{pmatrix} r \\ a_1^{\alpha_1} \ldots a_A^{\alpha_A} \end{pmatrix} = \frac{r!}{(a_1!)^{\alpha_1} \ldots (a_A!)^{\alpha_A}} \qquad \ldots \quad (2.8)$$

is the ordinary multinomial coefficient, and Σ' denotes summation over all unequal values of the suffices $i_1, \ldots, i_\alpha$ while Σ denotes summation over all *different* terms obtained as the suffices $1, \ldots, \alpha$ are replaced by any permutation of α unequal numbers between 1 and n. (The latter is the standard notation in symmetric function theory for any such *monomial* symmetric function.)

The notation $(a_1^{\alpha_1} \ldots a_A^{\alpha_A})$ for a partition of r may be called its "condensed form"; the same partition may be written in "extended form" as $(a_1 \ldots a_\alpha)$. For example, $(2^2 1)$ and (221) both represent the same partition of 5. Of course the meanings of $a_1, a_2, \ldots$ have been changed, but no confusion between the notations should arise, and the extended form often leads to simpler-looking (and simpler-printing) formulae. In extended form, (2.6) may be written

$$k_r = \sum_{(a)pr} (-)^{\alpha-1}(\alpha-1)! \begin{pmatrix} r \\ a_1 \ldots a_\alpha \end{pmatrix} \frac{\Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha}}{n^{(\alpha)}}$$

$$= \sum_{(a)pr} \begin{pmatrix} r \\ a_1 \ldots a_\alpha \end{pmatrix} B_{(k)}(a_1 \ldots a_\alpha)\, \Sigma\, x_1^{a_1} \ldots x_\alpha^{a_\alpha}, \qquad \ldots \quad (2.9)$$

where
$$B_{(k)}(a_1 \ldots a_\alpha) = \frac{(-)^{\alpha-1}(\alpha-1)!}{n^{(\alpha)}}. \qquad \ldots \quad (2.10)$$

It should be remembered that Σ in (2.9) still denotes a *monomial* symmetric function.

$B_{(k)}(a_1 \ldots a_\alpha)$ is thus the coefficient of $x_1^{a_1} \ldots x_\alpha^{a_\alpha}$ in k_r, reduced by the appropriate multinomial coefficient ; actually it is a function only of α, the number of parts in the partition, but the corresponding quantities for other sets of statistics than the k_r depend in more detail on the partitions, as will be seen immediately.

The sample estimates of μ_r' and μ_r, as usually defined, are $m_r' = n^{-1}\Sigma\, x_\gamma^r$ and

$$m_r = n^{-1}\Sigma(x_\gamma - m_1')^r = \sum_{j=0}^{r-2} (-)^j \binom{r}{j} m'_{r-j}\, m_1'^{\,j} + (-)^{r-1}(r-1) m_1'^{\,r}. \qquad \ldots \quad (2.11)$$

We may write these formulae in the form

$$m_r' = \sum_{(a)\mathrm{pr}} \binom{r}{a_1 \ldots a_\alpha} B_{(m')}(a_1 \ldots a_\alpha) \Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha}, \qquad \ldots \quad (2.12)$$

$$m_r = \sum_{(a)\mathrm{pr}} \binom{r}{a_1 \ldots a_\alpha} B_{(m)}(a_1 \ldots a_\alpha) \Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha}, \qquad \ldots \quad (2.13)$$

where
$$B_{(m')}(r) = 1/n,\ B_{(m')}(a_1 \ldots a_\alpha) = 0 \quad (\alpha \geqslant 2) ; \qquad \ldots \quad (2.14)$$

$$B_{(m)}(2) = (n-1)/n^2,\ B_{(m)}(1^2) = -1/n^2 ; \qquad \ldots \quad (2.15)$$

$$B_{(m)}(3) = (n-1)(n-2)/n^3,\ B_{(m)}(21) = -(n-2)/n^3,\ B_{(m)}(1^3) = 2/n^3 ; \qquad \ldots \quad (2.16)$$

$$\begin{aligned} &B_{(m)}(4) = (n-1)(n^2-3n+3)/n^4,\ B_{(m)}(31) = -(n^2-3n+3)/n^4, \qquad \ldots \quad (2.17)\\ &B_{(m)}(2^2) = (2n-3)/n^4,\ B_{(m)}(21^2) = (n-3)/n^4,\ B_{(m)}(1^4) = -3/n^4 ; \end{aligned}$$

$$\begin{aligned} &B_{(m)}(5) = (n-1)(n-2)(n^2-2n+2)/n^5,\ B_{(m)}(41) = -(n-2)(n^2-2n+2)/n^5, \ldots \quad (2.18)\\ &B_{(m)}(32) = (n-2)^2/n^5,\ B_{(m)}(31^2) = (n^2-3n+4)/n^5,\\ &B_{(m)}(2^21) = -2(n-2)/n^5,\ B_{(m)}(21^3) = -(n-4)/n^5,\ B_{(m)}(1^5) = 4/n^5 ; \end{aligned}$$

$$\begin{aligned} &B_{(m)}(6) = (n-1)(n^4-5n^3+10n^2-10n+5)/n^6, \qquad \ldots \quad (2.19)\\ &B_{(m)}(51) = -(n^4-5n^3+10n^2-10n+5)/n^6,\\ &B_{(m)}(42) = (n^3-4n^2+7n-5)/n^6,\ B_{(m)}(41^2) = (n^3-4n^2+6n-5)/n^6,\\ &B_{(m)}(3^2) = -(2n^2-6n+5)/n^6,\ B_{(m)}(321) = -(n^2-4n+5)/n^6,\\ &B_{(m)}(31^3) = -(n^2-3n+5)/n^6,\ B_{(m)}(2^3) = (3n-5)/n^6,\\ &B_{(m)}(2^21^2) = (2n-5)/n^6,\ B_{(m)}(21^4) = (n-5)/n^6,\ B_{(m)}(1^6) = -5/n^6 ; \end{aligned}$$

and so on.

It may be noted that while m_r' is an unbiased estimator of μ_r', and k_r of κ_r, m_r is not an unbiased estimator of μ_r. This accounts for the complexity of the $B_{(m)}$

coefficients.[1] For if we define a series of statistics l_r for which $El_r = \mu_r$, we have, from (2.1),

$$l_r = \sum_{j=0}^{r-2} (-)^j \binom{r}{r-j \quad 1^j} \frac{\Sigma x_1 \ldots x_j x_{j+1}^{r-j}}{n^{(j+1)}} + (-)^{r-1}(r-1) \binom{r}{1^r} \frac{\Sigma x_1 \ldots x_r}{n^{(r)}} \quad \ldots \quad (2.20)$$

$$= \sum_{(a)\text{p}r} \binom{r}{a_1 \ldots a_\alpha} B_{(l)}(a_1 \ldots a_\alpha) \Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha},$$

where

$$B_{(l)}(r-\alpha+1 \quad 1^{\alpha-1}) = (-)^{\alpha-1}/n^{(\alpha)} \qquad (\alpha = 1, \ldots, r-1),$$

$$B_{(l)}(1^r) = (r-1)(-)^{r-1}/n^{(r)}, \quad \ldots \quad (2.21)$$

while all $B_{(l)}$ for partitions containing more than one non-unit part vanish. For reference we note the following expressions for the l-statistics in terms of the sums of powers, $s_r = \Sigma x_\gamma^r$:

$$l_2 = [ns_2 - s_1^2]/n^{(2)} = \frac{n}{n-1} m_2, \quad \ldots \quad (2.22)$$

$$l_3 = [n^2 s_3 - 3ns_2 s_1 + 2s_1^3]/n^{(3)} = \frac{n^2}{(n-1)(n-2)} m_3, \quad \ldots \quad (2.23)$$

$$l_4 = [n(n^2-2n+3)s_4 - 4(n^2-2n+3)s_3 s_1 - 3(2n-3)s_2^2 + 6ns_2 s_1^2 - 3s_1^4]/n^{(4)}$$

$$= \frac{n}{(n-1)(n-2)(n-3)} [(n^2-2n+3)m_4 - 3(2n-3)m_2^2], \quad \ldots \quad (2.24)$$

$$l_5 = [n^2(n^2-5n+10)s_5 - 5n(n^2-5n+10)s_4 s_1 - 10n(n-2)s_3 s_2 + 10(n^2-4n+8)s_3 s_1^2$$
$$+30(n-2)s_2^2 s_1 - 10ns_2 s_1^3 + 4s_1^5]/n^{(5)}$$

$$= \frac{n^2}{(n-1)(n-2)(n-3)(n-4)} [(n^2-5n+10)m_5 - 10(n-2)m_3 m_2], \quad \ldots \quad (2.25)$$

$$l_6 = [n(n^4-9n^3+31n^2-39n+40)s_6 - 6(n^4-9n^3+31n^2-39n+40)s_5 s_1$$
$$-15(n^3-8n^2+29n-40)s_4 s_2 + 15(n^3-8n^2+23n-10)s_4 s_1^2 - 40(n^2-6n+10)s_3^2$$
$$+60(n^2-4n+5)s_3 s_2 s_1 - 20(n^2-6n+15)s_3 s_1^3 + 15(3n-10)s_2^3 - 45(2n-10)s_2^2 s_1^2$$
$$+15ns_2 s_1^4 - 5s_1^6]/n^{(6)}$$

$$= \frac{n}{(n-1)(n-2)(n-3)(n-4)n-5)} [(n^4-9n^3+31n^2-39n+40)m_6$$

$$-15(n^3-8n^2+29n-40)m_4 m_2 - 40(n^2-6n+10)m_3^2 + 15n(3n-10)m_2^3]. \quad \ldots \quad (2.26)$$

---

[1] However it ought to be mentioned that St. Georgescu's (1932) methods seem to work best for the m's as ordinarily defined.

In general, if we are dealing with an arbitrary set of z-statistics, z_r being a homogeneous polynomial symmetric function of the x's of degree r, then we shall denote the coefficient of

$$\begin{pmatrix} r \\ a_1 \ldots a_\alpha \end{pmatrix} x_1^{a_1} \ldots x_\alpha^{a_\alpha} \qquad \ldots \quad (2.27)$$

in the expansion of z_r by $B_{(z)}\,(a_1 \ldots a_\alpha)$.

Now for any such set of z-statistics there are three problems which can be solved by an extension of Fisher's combinatorial method :

(i) The expression of (ordinary or mixed) moments of the z_r in terms of population moments. We call this the $\mu' z \mu'$ problem (with $z = k,\ m',\ m,\ l,\ \ldots,$ as appropriate). The moments of the z_r are to be taken about zero; so are the population moments. But if the particular set of z-statistics considered are seminvariant[1] and z_1 is not involved, we may take the origin at the mean and so express the sampling moments in terms of moments about the mean.

(ii) The expression of (ordinary or mixed) moments of the z_r in terms of cumulants. This is the $\mu' z \kappa$ problem.

(iii) The expression of (ordinary or mixed) cumulants of the z_r in terms of cumulants. This is the $\kappa z \kappa$ problem.

The most concise results are those of the type $\kappa z \kappa$, but our proof proceeds through the stages $\mu' z \mu'$ and $\mu' z \kappa$. Occasionally one of these latter problems is of interest in itself. For example, in determining the constants of the distribution of $g_1 = k_3/k_2^{3/2}$ in sampling from a *normal* population, it is known that $Eg_1^{2r} = Ek_3^{2r}/Ek_2^{3r}$, whereas no such simple relation holds among the cumulants.

We shall use the notations $\mu'_{(z)}(r_1^{\rho_1} \ldots r_R^{\rho_R})$ or $\mu'_{(z)}(r_1 \ldots r_\rho)$ for the $(\rho_1; \ldots, \rho_R)$-th moment about the origin of $z_{r_1}, \ldots, z_{r_R}$, and $\kappa_{(z)}(r_1^{\rho_1} \ldots r_R^{\rho_R})$ for their $(\rho_1, \ldots, \rho_R)$-th cumulant. The number $w = \Sigma\, \rho_i r_i$ (or Σr_i in the extended notation) is called the *order* of the moment or cumulant.

We shall now state the rules applicable to the three problems. For the convenience of readers of Kendall's book (1947), the numbering and statements of these will be as far as possible in accordance with his, but since some of the rules do not apply to all these problems, at any rate without modification, we shall indicate before each statement, in parenthesis, its range of validity.

Rule 1 : $(\kappa z \kappa,\ \mu' z \kappa,\ \mu' z \mu')$. A moment or cumulant of z-statistics of order w consists of the sum of terms each of which is a multiple of a product of parent moments or cumulants of order w.

---

[1] In Kendall's sense; see section 1. Since z_1 is a homogeneous polynomial symmetric function of degree one, it is a multiple of the sample mean, and cannot possibly be seminvariant (unless it is identically zero). When, therefore, we say that the z_r are seminvariant, we always implicitly exclude z_1. The k's, m's and l's are seminvariant statistics.

Rule 2 : ($\kappa z \kappa$). *If the z-statistics are seminvariant,* no term in $\kappa_{(z)}(r_1 \ldots r_\rho)$ contains κ_1 except $\kappa_{(z)}(1)$ itself.

Rule 2a : ($\mu' z \kappa$). *If the z-statistics are seminvariant,* no term in $\mu'_{(z)}(r_1 \ldots r_\rho)$ contains κ_1, provided that z_1 is not involved in the formula.

Rule 2b : ($\mu' z \mu'$). Strictly speaking there is no corresponding rule in this case, but, as remarked above, if z_1 is not involved we may take the origin at the population mean, so that $\mu'_1 = 0$, $\mu'_r = \mu_r$.

To find the coefficient of $\kappa_{\ 1} \ldots \kappa_{t_\tau}$ in $\kappa_{(z)}(r_1 \ldots r_\rho)$; we consider all different *bipartitions* or *arrays*

$$\begin{array}{c|c} & t_1 \\ & t_2 \\ & \vdots \\ & t_\tau \\ \hline r_1\, r_2\, \ldots\, r_\rho & w \end{array} \qquad \ldots \quad (2.28)$$

the body of the array being filled up by numbers in such a manner that the correct column and row totals result. The column and row totals are to be written in a fixed order. The same procedure is followed in the $\mu' z \kappa$ and $\mu' z \mu'$ problems.

Rule 3 : ($\kappa z \kappa$). Every array in which the numbers in the body of the array fall into two or more blocks, each confined to separate rows and columns, is to be ignored.

Rule 3a : ($\mu' z \kappa$). These dissectable arrays are not to be neglected. But use can be made of the fact that the κ-type pattern function of such an array (see Rules 4 and 6) is equal to the product of the κ-type pattern functions of the separate blocks (by Lemma 6.1).

Rule 3b : ($\mu' z \mu'$). The dissectable arrays are not to be neglected; neither is the above product rule true.

Rule 4 : ($\kappa z \kappa$, $\mu' z \kappa$, $\mu' z \mu'$). Subject to any ignoration of terms enjoined by Rule 3, to the coefficient of $\kappa_{t_1} \ldots \kappa_{t_\tau}$ in $\kappa(r_1 \ldots r_\rho)$ there will be a contribution corresponding to each way of completing the array (2.28). Such of these as do not vanish are composed of a *numerical coefficient* multiplied by a function of n called the *pattern function.* Similar remarks apply in the $\mu' z \kappa$ and $\mu' z \mu'$ problems.

Rule 5 : ($\kappa z \kappa$, $\mu' z \kappa$, $\mu' z \mu'$). The numerical coefficient is the number of ways in which the column totals, considered as composed of distinct individuals, can be allocated to form the array concerned, divided by $\tau_1! \ldots \tau_T!$. ($\tau_1, \ldots, \tau_T$ it will be remembered, are the sizes of the groups of equal t's.) For example the array

$$\begin{array}{ccc|c} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \qquad \ldots \quad (2.29)$$

has numerical coefficient

$$\binom{4}{21^2}\binom{4}{31}\binom{2}{1^2}\frac{1}{2!}=48. \qquad \dots \quad (2.30)$$

With it we may consider the 3 other arrays obtainable from it by interchanging the first two columns and the last two rows,[1] making a total coefficient of 192.

Rule 6 : $(\kappa z \kappa,\ \mu' z \kappa)$. The function of n, called the *κ-type pattern function*, depends, when k-statistics are under consideration, only on the pattern of non-zero elements in the array, not on their actual magnitudes. This is no longer true for general z-statistics, but we shall nevertheless retain the old name. The function is given by considering the *separations* of the rows into distinct groups or *separates*. For example a three-rowed array like (2.29) possesses five separations : one in which all three rows comprise one group or separate, three in which there is one separate of two rows and one of one row, and the *complete* separation in which there are three separates of one row each. The *specifications* of these separations, that is to say the arrays obtained by adding together element-by-element the rows of each separate, are respectively

$$\begin{array}[b]{ccc|c} 4 & 4 & 2 & 10 \\ \hline 4 & 4 & 2 & 10 \end{array} \quad \begin{array}[b]{ccc|c} 3 & 4 & 1 & 8 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \quad \begin{array}[b]{ccc|c} 3 & 3 & 2 & 8 \\ 1 & 1 & . & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} \quad \begin{array}[b]{ccc|c} 2 & 3 & 1 & 6 \\ 2 & 1 & 1 & 4 \\ \hline 4 & 4 & 2 & 10 \end{array} \quad \begin{array}[b]{ccc|c} 2 & 3 & 1 & 6 \\ 1 & 1 & . & 2 \\ 1 & . & 1 & 2 \\ \hline 4 & 4 & 2 & 10 \end{array} . \qquad \dots \quad (2.31)$$

(i) If the specification of a particular separation has χ rows, we associate with it the factor $n^{(\chi)}$.

(ii) We associate with a column $(b_1 \dots b_\beta)$ the factor $B_{(z)}(b_1 \dots b_\beta)$ defined by (2.27).

(iii) This is done for each column.

(iv) The factors given by (i), (ii) and (iii) are multiplied together for each separation and the results added to give the pattern function. Thus in our example the pattern function (for k-statistics) is

$$n\left(\frac{1}{n}\right)^3+2n^{(2)}\left(\frac{-1!}{n^{(2)}}\right)^2\frac{1}{n}+n^{(2)}\left(\frac{-1!}{n^{(2)}}\right)^3+n^{(3)}\frac{2!}{n^{(3)}}\left(\frac{-1!}{n^{(2)}}\right)^2=\frac{1}{(n-1)^2}. \qquad \dots \quad (2.32)$$

Rule 6a : $(\mu' z \mu')$. The function of n, called the *μ'-type pattern function*, is evaluated in the same way as the κ-type pattern function, except that only the complete separation (the original array) is considered. Thus if the array has τ rows and a typical column is $(a_1 \dots a_\alpha)$ then the μ'-type pattern function is obtained by multiplying $n^{(\tau)}$ by a factor $B_{(z)}(a_1 \dots a_\alpha)$ for each column. In our example this gives

$$n^{(3)}\ \frac{2!}{n^{(3)}}\left(\frac{-1!}{n^{(2)}}\right)^2=\frac{2}{n^2(n-1)^2}. \qquad \dots \quad (2.33)$$

---

[1] Care must be observed when using this short-cut that all the new arrays produced are actually different.

Bureau Ednl. [illegible] Research
DAVID H[illegible] COLLEGE
Dated

Rule 7 : ($\kappa k \kappa$, $\mu' k \kappa$). Any array containing a row which consists of a single non-zero entry, not forming a block by itself, has vanishing pattern function and can be neglected.

This rule is not true for the $\mu' k \mu'$ problem or for general z-statistics.

Rule 8 : ($\kappa z \kappa$, $\mu' z \kappa$, $\mu' z \mu'$). Any array containing a column which consists of a single entry r, not forming a block by itself, has a pattern function $B_{(z)}(r)$ times that of the array obtained by omitting that column. (For k-, m'- and l-statistics, $B_{(z)}(r) = n^{-1}$, but not for the m's).

Rule 9 : ($\kappa k \kappa$, $\mu' k \kappa$). Any array, the non-zero elements of which consist of two groups, unconnected except by precisely one column, has vanishing pattern function and can be neglected.

Like Rule 7, which is a particular case, *this rule is not true for the $\mu' k \mu'$ problem, or for general z-statistics.*

Rule 10 : ($\kappa z \kappa$). The expression for any $\kappa(r_1 \ldots r_\rho)$ which involves a unit part may be obtained from that without the part by (i) increasing the suffix of one of the κ's by unity in every possible way, and (ii) multiplying through by $B_{(z)}(1)$. (For $z_1 = k_1$ or m'_1, $B_{(z)}(1) = n^{-1}$.)

Rule 10a : ($\mu' z \kappa$). As above, but in addition we must add on $nB_{(z)}(1)\ \kappa_1$ times the original expression. (For $z_1 = k_1$ or m'_1, $nB_{(z)}(1)\ \kappa_1 = \kappa_1$.)

For example, from

$$\kappa_{(k)}(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1} \qquad \ldots \quad (2.34)$$

we derive

$$\kappa_{(k)}(2^2 1) = \frac{\kappa_5}{n^2} + \frac{4\kappa_3\kappa_2}{n(n-1)} \text{ and } \kappa_{(k)}(2^2 1^2) = \frac{\kappa_6}{n^3} + \frac{4\kappa_4\kappa_2}{n^2(n-1)} + \frac{4\kappa_3^2}{n^2(n-1)}, \qquad \ldots \quad (2.35)$$

and from

$$\mu'_{(k)}(2^2) = \frac{\kappa_4}{n} + \frac{n+1}{n-1}\kappa_2^2 \qquad \ldots \quad (2.36)$$

we derive

$$\mu'_{(k)}(2^2 1) = \frac{\kappa_5}{n^2} + \frac{\kappa_4\kappa_1}{n} + \frac{2(n+1)}{n(n-1)}\kappa_3\kappa_2 + \frac{n+1}{n-1}\kappa_2^2\kappa_1. \qquad \ldots \quad (2.37)$$

There is no corresponding rule for the $\mu' z \mu'$ problem.

Rules 1, 3, 3a, 3b, 4, 5, 6, 6a are the fundamental ones of the process; the others are merely auxiliary, and enable short-cuts to be made.

Finally we remark that the above rules are rather more general than the above account might at first sight suggest. Thus, because l_1 vanishes identically we might wish to consider the system $z_1 = m'_1$, $z_r = l_r (r \geqslant 2)$. This is covered by the rules. We could even evaluate such expressions as the covariance of k_4 and l_4, if the need arose. (The difficulty that there are two z-statistics of the same order is a purely notational one.)

3. Multivariate extension

Suppose now that each individual of our infinite population has p measurements made on it, denoted by $x^1, \dots, x^p$ and denote the sample of size n by $(x^1_\gamma, \dots, x^p_\gamma)$ $(\gamma = 1, \dots, n)$. The $(r_1, \dots, r_p)$-th population moment about the origin is $\mu'_{r_1 \dots r_p} = E(x^1)^{r_1} \dots (x^p)^{r_p}$, and the corresponding cumulants are defined by formal relation

$$\sum \frac{\kappa_{r_1 \cdots r_p} \theta_1^{r_1} \dots \theta_p^{r_p}}{r_1! \dots r_p!} = \log \sum \frac{\mu'_{r_1 \cdots r_p} \theta_1^{r_1} \dots \theta_p^{r_p}}{r_1! \dots r_p!} . \qquad \dots \quad (3.1)$$

Thus

$$\kappa_{r_1 \cdots r_p} = \sum_{(a)\mathrm{p}(r)} (-)^{\alpha-1}(\alpha-1)! \left[\begin{matrix} (r_1 \cdots r_p) \\ (a_{11} \dots a_{p1})^{\alpha_1} \dots (a_{1A} \dots a_{pA})_p{}^{\alpha_A} \end{matrix}\right] \mu'^{\alpha_1}_{a_{11} \cdots a_{p1}} \cdots \mu'^{\alpha_A}_{a_{1A} \cdots a_{pA}}, \qquad \dots \quad (3.2)$$

where

$$\left[\begin{matrix} (r_1 \dots r_p) \\ (a_{11} \dots a_{p1})^{\alpha_1} \dots (a_{1A} \dots a_{pA})^{\alpha_A} \end{matrix}\right] = \frac{r_1! \dots r_p!}{(a_{11}! \dots a_{p1}!)^{\alpha_1} \alpha_1! \dots (a_{1A}! \dots a_{pA}!)^{\alpha_A} \alpha_A!} \qquad \dots \quad (3.3)$$

is a "reduced multinomial coefficient" and $\sum\limits_{(a)\mathrm{p}(r)}$ denotes the sum over all partitions $((a_{11} \dots a_{p1})^{\alpha_1} \dots (a_{1A} \dots a_{pA})^{\alpha_A})$ of the multipartite number $(r_1 \dots r_p)$, with $\Sigma \alpha_j a_{ij} = r_i$ $(i = 1, \dots, p)$, $\Sigma \alpha_j = \alpha$.

The corresponding k-statistic, defined to have expectation $\kappa_{r_1 \cdots r_p}$, is

$$k_{r_1 \cdots r_p} = \sum_{(a)\mathrm{p}(r)} (-)^{\alpha-1}(\alpha-1)! \left(\begin{matrix} (r_1 \dots r_p) \\ (a_{11} \dots a_{p1})^{\alpha_1} \dots (a_{1A} \dots a_{pA})^{\alpha_A} \end{matrix}\right) \times$$

$$\times \frac{\Sigma (x^1_1 \dots x^1_{\alpha_1})^{\alpha_{11}} \dots (x^p_1 \dots x^p_{\alpha_1})^{\alpha_{p1}} \dots (x^1_{\alpha-\alpha_A+1} \cdots x^1_\alpha)^{\alpha_{1A}} \dots (x^p_{\alpha-\alpha_A+1} \cdots x^p_\alpha)^{\alpha_{pA}}}{n^{(\alpha)}} . \qquad \dots (3.4)$$

Or, in extended notation,

$$k_{r_1 \cdots r_p} = \sum_{(a)\mathrm{p}(r)} \left(\begin{matrix} (r_1 \dots r_p) \\ (a_{11} \dots a_{p1}) \dots (a_{1\alpha} \dots a_{p\alpha}) \end{matrix}\right) B_{(k)}((a_{11} \dots a_{p1}) \dots (a_{1\alpha} \dots a_{p\alpha})) \times$$

$$\times \Sigma\, (x^1_1)^{a_{11}} \dots (x^p_1)^{a_{p1}} \dots (x^1_\alpha)^{a_{1\alpha}} \dots (x^p_\alpha)^{a_{p\alpha}}, \qquad \dots \quad (3.5)$$

where

$$B_{(k)}(\dots) = \frac{(-)^{\alpha-1}(\alpha-1)!}{n^{(\alpha)}} , \qquad \dots \quad (3.6)$$

as in the univariate case. In each of (3.4) and (3.5) the expression in large round brackets denotes an ordinary multinomial coefficient ((3.3) without the factors $a_1! \dots \alpha_A!$ in the denominator) and Σ denotes "symmetric function" summation, extending to all different terms, as in equations (2.6) and (2.9).

Other sets of statistics than the k's may be defined, as in section 2. This will be considered in more detail later in the section. For a general set of z-statistics the quantity $B_{(z)}\,(\dots)$ is defined as the coefficient of

$$\begin{pmatrix} (r_1 \dots r_p) \\ (a_{11} \dots a_{p1})\dots(a_{1\alpha}\dots a_{p\alpha}) \end{pmatrix} (x_1^1)^{a_{11}}\dots(x_1^p)^{a_{p1}}\dots(x_\alpha^1)^{a_{1\alpha}}\dots(x_\alpha^p)^{a_{p\alpha}} \qquad \dots \quad (3.7)$$

in $z_{r_1..r_p}$. We shall write $\mu'_{(z)}((r_{11}\dots r_{p1})^{\rho_1} \dots (r_{1R}\dots r_{pR})^{\rho}\ )$ or $\mu'_{(z)}((r_{11}\dots r_{p1})\dots(r_{1\rho}\dots r_{p\rho}))$ for the $(\rho_1, \dots, \rho_R)$-th moment of $z_{r_{11}\dots r_{p1}}, \dots, z_{r_{1R}..r_{pR}}$, and similarly for their $(\rho_1, \dots, \rho_R)$-th cumulant. The two forms correspond to the condensed and extended notations for the partition, $((r_{11} \dots r_{p1})^{\rho_1} \dots (r_{1R} \dots r_{pR})^{\rho_R}) = ((r_{11} \dots r_{p1})\dots(r_{1\rho} \dots r_{p\rho}))$, of the multipartite number $(w_1 \dots w_p)$ which represents the *order* (in the p variates) of the moment or cumulant.

The rules for the multivariate case may be stated as follows.

Rule 1 : $(\kappa z \kappa,\ \mu' z \kappa,\ \mu' z \mu')$. A moment or cumulant of z-statistics of order $(w_1 \dots w_p)$ consists of the sum of terms each of which is a product of parent moments or cumulants of order $(w_1 \dots w_p)$.

Rule 2 : $(\ \kappa z \kappa\ )$. *If the z-statistics are seminvariant*, no term in $\kappa_{(z)}((r_{11} \dots r_{p1}) \dots (r_{1\rho} \dots r_{p\rho}))$ contains $\kappa_{10\cdots0}$ except $\kappa_{(z)}((10 \dots 0))$ itself (and similarly for the other first-order cumulants).

Rule 2a : $(\mu' z \kappa)$. *If the z-statistics are seminvariant*, no term in $\mu'_{(z)}((r_{11}\dots r_{p1}) \dots (r_{1\rho} \dots r_{p\rho}))$ contains $\kappa_{10\cdots0}$, provided that $z_{10\cdots0}$ is not involved in the formula (and similarly for the other first order cumulants).

To find the coefficient of $\kappa_{t_{11}\dots t_{p1}} \cdots \kappa_{t_{1\tau}\dots t_{p\tau}}$ in $\kappa_{(z)}((r_{11} \dots r_{p1}) \dots (r_{1\rho} \dots r_{p\rho}))$, we consider all different arrays

$$\begin{array}{c|l} & (t_{11} \dots t_{p1}) \\ & \quad\dots \\ & (t_{1\tau} \dots t_{p\tau}) \\ \hline (r_{11} \dots r_{p1}) \dots (r_{1\rho} \dots r_{p\rho}) & (w_1 \dots w_p) \end{array} \qquad \dots \quad (3.8)$$

the body of the array being filled in by p-partite numbers in such a way that correct column and row totals result for each of the p parts. The column and row totals are to be written in a fixed order. The same procedure is followed in the $\mu' z \kappa$ and $\mu' z \mu'$ problems.

Rule 3 : (Arrays to be neglected, etc.) As in the univariate case, *Rule* 4 : (Form of coefficient : numerical coefficient and pattern function). As in the univariate case, with trivial modifications.

Rule 5 : (Numerical coefficient). As in the univariate case. For example, the array

$$\begin{array}{ccc|c} (20) & (12) & (10) & (42) \\ (01) & (01) & & (02) \\ (01) & & (10) & (11) \\ \hline (22) & (13) & (20) & (55) \end{array} \qquad \dots \quad (3.9)$$

has numerical coefficient

$$\binom{2}{2}\binom{2}{1^2}\binom{1}{1}\binom{3}{21}\binom{2}{1^2} = 12. \qquad \dots \quad (3.10)$$

This array may be compared with (2.29) : if the two parts of each bipartite number are added we recover (2.29). But, since the row totals (02) and (11) are not now regarded as equal, we do not divide by 2! as we did in (2.30). Likewise there are no further arrays to be considered at the same time as (3.9).

Rule 6 : (Pattern functions). As in the univariate case, with trivial modifications. Thus the κ-type pattern function (for k-statistics) of (3.9) is the same as that of (2.29), namely $1/(n-1)^2$.

Rule 7 : (Row with single non-zero element).

Rule 8 : (Column with single non-zero element).

Rule 9 : (Two groups joined by a single column). These rules are the same as in the univariate case.

Rule 10 : ($\kappa z \kappa$). The expression for any $\kappa_{(z)}((r_{11} \dots r_{p1})\dots(r_{1\rho} \dots r_{p\rho}))$ which involves a unit part $(10 \dots 0)$ may be obtained from that without the part by (i) increasing the appropriate suffix of one of the κ's by unity in every possible way, and (ii) multiplying through by $B_{(z)}((10 \dots 0))$. A similar rule holds for adding any other unit part.

Rule 10a : ($\mu' z \kappa$). As above, but in addition we must add on $nB_{(z)}((10 \dots 0))\ \kappa_{10 \cdots 0}$ times the original expression.

For example, from

$$\kappa_{(k)}((20)(11)) = \frac{\kappa_{31}}{n} + \frac{2\kappa_{20}\kappa_{11}}{n-1} \qquad \dots \quad (3.11)$$

we derive
$$\kappa_{(k)}((20)(11)(10)) = \frac{\kappa_{41}}{n^2} + \frac{2(\kappa_{30}\kappa_{11} + \kappa_{20}\kappa_{21})}{n(n-1)} \qquad \dots \quad (3.12)$$

and
$$\kappa_{(k)}((20)(11)(01)) = \frac{\kappa_{32}}{n^2} + \frac{2(\kappa_{21}\kappa_{11} + \kappa_{20}\kappa_{12})}{n(n-1)}, \qquad \dots \quad (3.13)$$

and from
$$\mu'_{(k)}((20)(11)) = \frac{\kappa_{31}}{n}+\frac{n+1}{n-1}\,\kappa_{20}\kappa_{11} \qquad \dots \quad (3.14)$$

we find
$$\mu'_{(k)}((20)(11)(01)) = \frac{\kappa_{32}}{n^2}+\frac{\kappa_{31}\kappa_{01}}{n}+\frac{n+1}{n(n-1)}(\kappa_{21}\kappa_{11}+\kappa_{20}\kappa_{12})+\frac{n+1}{n-1}\kappa_{20}\kappa_{11}\kappa_{01}$$
$$\dots \quad (3.15)$$

It is of particular interest to consider the results we obtain when the number of potentially distinct variates involved in a formula is equal to its total order, $w=w_1+\ldots+w_p$. Each variate is then involved linearly in the formula, so for brevity we shall call this the multilinear case. By *potentially* distinct variates we mean that in all our work on the multilinear case we shall assume that the variates x^i, x^j,... are distinct for enumerative purposes; but this will not stop us from equating some of them to give univariate or less general multivariate results at a later stage. As a simple illustration, if we have two variates x and y, then the monomial symmetric function $\Sigma x_1 y_2$ contains $n(n-1)$ terms. If we now consider the particular case when $x \equiv y$, $\Sigma x_1 y_2$ becomes $2\Sigma x_1 x_2$; for $\Sigma x_1 x_2$ only has $\frac{1}{2}n(n-1)$ terms, but $\Sigma x_1 y_2$ refers still to two potentially (though not actually) distinct variates and still connotes a sum of $n(n-1)$ terms.

Kaplan's "tensor" notation is useful in the multilinear case. We consider an unlimited (or sufficiently large) set of variates x^1, x^2, ..., all potentially distinct. The tensor notation for moments is exemplified by

$$\mu'^{ijkl} = Ex^i x^j x^k x^l \qquad \dots \quad (3.16)$$

for a (multilinear) fourth moment. For example μ'^{1111} is μ'_4 (or $\mu'_{400\dots}$) in the 'power' notation, μ'^{1123} is μ'_{211}, and so on. Similar notation is used for the cumulants, so that it can be seen that (3.1) becomes

$$\frac{\kappa^i\,\theta_i}{1!}+\frac{\kappa^{ij}\theta_i\theta_j}{2!}+\ldots = \log\left[\ 1+\ \frac{\mu'^i\theta_i}{1!}\ +\frac{\mu'^{ij}\theta_i\theta_j}{2!}+\ldots\ \right], \qquad \dots \quad (3.17)$$

where use is made of the summation convention for repeated upper and lower indices (e.g. $\kappa^{ij}\theta_i\theta_j = \Sigma_{ij}\,\kappa^{ij}\theta_i\theta_j$). Particular cases of (3.2) are

$$\kappa^i = \mu'^i,$$

$$\kappa^{ij} = \mu'^{ij}-1!\,\mu'^i\mu'^j \qquad \dots \quad (3.18)$$

$$\dots \quad (3.19)$$
$$\kappa^{ijk} = \mu'^{ijk}-1!\,(\mu'^{ij}\mu'^k+\mu'^{ik}\mu'^j+\mu'^{jk}\mu'^i)+2!\,\mu'^i\mu'^j\mu'^k,$$

$$\dots \quad (3.20)$$
$$\kappa^{ijkl} = \mu'^{ijkl}-1!\,\overset{4}{\Sigma}\,\mu'^{ijk}\mu'^l-1!\,\overset{3}{\Sigma}\,\mu'^{ij}\mu'^{kl}+2!\,\overset{6}{\Sigma}\,\mu'^{ij}\mu'^k\mu'^l-3!\mu'^i\mu'^j\mu'^k\mu'^l, \qquad \dots \quad (3.21)$$

where for example,

$$\overset{3}{\Sigma}\,\mu'^{ij}\mu'^{kl} = \mu'^{ij}\mu'^{kl}+\mu'^{ik}\mu'^{jl}+\mu'^{il}\mu'^{jk},$$
$$\dots \quad (3.22)$$

the sum being over all $\left[\begin{smallmatrix}4\\2^2\end{smallmatrix}\right] = 3$ distinct terms. In general (3.2) becomes

$$\kappa^{i_1 \cdots\, i_r} = \sum_{(a)\mathrm{p}r} (-)^{\alpha-1}\ (\alpha-1)!\ \Sigma\ \mu'^{\,i_1\cdots i_{a_1}} \ldots \mu'^{\,i_{r-a_\alpha+1}\cdots i_r} \qquad \ldots \quad (3.23)$$

the first sum being over all partitions $(a_1 \ldots a_\alpha)$ of the ordinary number r, and the second over all distinct terms obtained by permuting the indices $i_1, \ldots, i_r$. The companion equation giving moments in terms of cumulants is simply

$$\mu'^{i_1 \ldots i_r} = \sum_{(a)\mathrm{p}r}\ \Sigma\ \kappa^{i_1 \ldots i_{a_1}} \ldots \kappa^{i_{r-a_\alpha+1} \cdots i_r}. \qquad \ldots \quad (3\ 24)$$

Equations (3.4) and (3.5) become

$$k^{i_1 \cdots i_r} = \sum_{(a)\mathrm{p}r} (-)^{\alpha-1} (\alpha-1)!\ \frac{\Sigma\ \Sigma\ x_1^{i_1} \ldots x_1^{i_{a_1}} \ldots x_\alpha^{i_{r-a_\alpha+1}} \ldots x_\alpha^{i_r}}{n^{(\alpha)}}$$

$$= \sum_{(a)\mathrm{p}r} B_{(k)}\,(a_1 \ldots a_\alpha)\ \Sigma\,\Sigma\ x_1^{i_1} \ \ldots\ x_1^{i_{a_1}} \ldots x_\alpha^{i_{r-a_\alpha+1}} \ldots x_\alpha^{i_r}. \qquad \ldots \quad (3.25)$$

where
$$B_{(k)}(a_1 \ldots a_\alpha) = \frac{(-)^{\alpha-1}(\alpha-1)!}{n^{(\alpha)}}. \qquad \ldots \quad (3.26)$$

The third summation in (3.25) is over all different terms[1] obtained by replacing the suffices $1, \ldots, \alpha$ by any other set of α different numbers between 1 and n, while the second is over all terms, obtained by permuting $i_1, \ldots, i_r$, which are distinct (after the third summation has been carried out). For example,

$$k^{ijkl} = \frac{1}{n} \sum x_1^i\, x_1^j\, x_1^k\, x_1^l - \frac{1!}{n^{(2)}} \sum^{4} \sum x_1^i\, x_1^j\, x_1^k\, x_2^l - \frac{1!}{n^{(2)}} \sum^{3} \sum x_1^i\, x_1^j\, x_2^k\, x_2^l$$

$$+ \frac{2!}{n^{(3)}} \sum^{6} \sum x_1^i\, x_1^j\, x_2^k\, x_3^l - \frac{3!}{n^{(4)}} \sum x_1^i\, x_2^j\, x_3^k\, x_4^l. \qquad \ldots \quad (3.27)$$

It is of great importance to notice that if appropriate variates are equated in equations (3.17) to (3.24) we recover the univariate results (2.2) to (2.5) and the multivariate results (3.1) and (3.2). Since k-statistics are unbiased estimators of the corresponding κ's, it follows that (3.25) includes (2.9) and (3.5) in a similar way. It follows that, as far as k-statistics are concerned, any formula established in the multilinear case remains true when some or all of the variates are equated. Since many of the results are far easier to prove rigorously in the multilinear case, this is a great convenience. But in order to ensure that this holds for any arbitrary set of

---

[1] There are now just $n^{(\alpha)}$ such terms, since $x^{i_1}, \ldots, x^{i_r}$ are potentially distinct.

z-statistics, it is necessary to define the multilinear z's in an appropriate fashion. If, as usual, z_r is given by

$$z_r = \sum_{(a)\mathrm{p}r} \begin{pmatrix} r \\ a_1 \dots a_\alpha \end{pmatrix} B_{(z)}\,(a_1 \dots a_\alpha)\, \Sigma\, x_1^{a_1} \dots x_\alpha^{a_\alpha}, \qquad \dots \quad (3.28)$$

then it is an unbiased estimator of

$$\zeta_r = \sum_{(a)\mathrm{p}r} \begin{bmatrix} r \\ a_1 \dots a_\alpha \end{bmatrix} n^{(\alpha)} B_{(z)}\,(a_1 \dots a_\alpha)\, \mu'_{a_1} \dots \mu'_{a_\alpha}. \qquad \dots \quad (3.29)$$

Any reasonable multilinear extension of (3.29) must be symmetrical in the r variates, and with this proviso the only possible generalization is

$$\zeta^{i_1 \dots i_r} = \sum_{(a)\mathrm{p}r} n^{(\alpha)} B_{(z)}\,(a_1 \dots a_\alpha)\, \Sigma\, \mu'^{i_1 \dots i_{a_1}} \dots \mu'^{i_{r-a_\alpha+1} \dots i_r}, \qquad \dots \quad (3.30)$$

interpreted like (3.23). On equating $x^{i_1}, \dots, x^{i_r}$ we regain (3.29), and by equating the variates in groups we would obtain the less general multivariate ζ-parameters. Finally the unbiased estimator of (3.30) is

$$z^{i_1 \dots i_r} = \sum_{(a)\mathrm{p}r} B_{(z)}\,(a_1 \dots a_\alpha)\, \Sigma\, \Sigma x_1^{i_1} \dots x_1^{i_{a_1}} \dots x_\alpha^{i_{r-a_\alpha+1}} \dots x_\alpha^{i_r}, \qquad \dots \quad (3.31)$$

interpreted like (3.25). This reduces to (3.28) on equating all the variates, so that once multilinear results using this definition of the z's have been proved, they can be particularized to give univariate, bivariate[1], ..., results. It may be noted that the definition (3.31) is completely natural as far as the special statistics we have considered are concerned; thus it gives rise to the multilinear statistics

$$m'^{i_1 \dots i_r} = n^{-1}\, \Sigma x_\gamma^{i_1} \dots x_\gamma^{i_r} \qquad \dots \quad (3.32)$$

and

$$m^{i_1 \dots i_r} = n^{-1}\, \Sigma (x_\gamma^{i_1} - m'^{i_1}) \dots (x_\gamma^{i_r} - m'^{i_r}) \qquad \dots \quad (3.33)$$

while $k^{i_1 \dots i_r}$ and $l^{i_1 \dots i_r}$ are unbiased estimates of $\kappa^{i_1 \dots i_r}$ and of

$$\mu^{i_1 \dots i_r} = E(x^{i_1} - \mu'^{i_1}) \dots (x^{i_r} - \mu'^{i_r}). \qquad \dots \quad (3.34)$$

The $B_{(m')}$, $B_{(m)}$ and $B_{(l)}$ are given in section 2, while the l's are given in terms of sums of products by equations (2.21) to (2.26) on replacing

$$\begin{bmatrix} r \\ a_1 \dots a_\alpha \end{bmatrix} s_{a_1} \dots s_{a_\alpha} \text{ by } \Sigma\, s^{i_1 \dots i_{a_1}} \dots s^{i_{r-a_\alpha+1} \dots i_r}, \qquad \dots \quad (3.35)$$

---

[1] The bivariate and higher results will, of course, be for z-statistics which are symmetrically defined in the sense that the coefficients $B_{(z)}$ are unaffected by a permutation of variates applied to the whole equation. In the contrary case (which is most unlikely to occur in practice) the coefficient $B_{(z)}$ in (3.30) and (3.31) would have to be made to depend on all the indices $i_1 \dots i_{a_1}, \dots, i_{r-a_\alpha+1} \dots i_r$, and not just on $a_1, \dots, a_\alpha$.

where, for example, $s^{ijkl} = \Sigma x_\gamma^i x_\gamma^j x_\gamma^k x_\gamma^l$. Thus (2.24) becomes

$$l^{ijkl} = [n(n^2-2n+3)s^{ijkl} - (n^2-2n+3)\overset{4}{\Sigma}\, s^{ijk}s^l - (2n-3)\overset{3}{\Sigma}\, s^{ij}s^{kl}$$

$$+\, n\overset{6}{\Sigma}\, s^{ij}s^k s^l - 3s^i s^j s^k s^l]/n^{(4)}$$

$$= \frac{n}{(n-1)(n-2)(n-3)}\,[(n^2-2n+3)m^{ijkl}-(2n-3)\overset{3}{\Sigma}\, m^{ij}m^{kl}]. \qquad \ldots \quad (3.36)$$

4. PROOF OF THE FUNDAMENTAL RULES FOR THE $\mu' z \mu'$ PROBLEM

Rule 1 (for all the problems) follows from elementary considerations, and a formal proof will be omitted.

Consider the product $z_{r_1} \ldots z_{r_\rho}$ corresponding to the partition $(r_1 \ldots r_\rho)$ of the number $w = \Sigma r_h$. Each of the factors may be expressed in the form given by (2.27), or

$$z_{r_h} = \left[\sum_{(a)\mathrm{pr}} \begin{pmatrix} r \\ a_1 \ldots a_\alpha \end{pmatrix} B_{(z)}(a_1 \ldots a_\alpha)\, \Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha}\right]_h, \qquad \ldots \quad (4.1)$$

the notation $[\ ]_h$ being a helpful abbreviation indicating that the letters r, a and α should really possess an extra suffix h to show which r is being partitioned. Diagrammatically we may represent $z_{r_1} \ldots z_{r_\rho}$ by the partition $(r_1 \ldots r_\rho)$ written horizontally, and each of the r's yields sub-partitions (written vertically) corresponding to the terms of (4.1).

For example,

$$k_2^2 = \left(\frac{\Sigma\, x_1^2}{n} - 2\,\frac{\Sigma\, x_1x_2}{n^{(2)}}\right)^2$$

$$= \frac{(\Sigma\, x_1^2)^2}{n^2} - 2\,\frac{\Sigma\, x_1^2\,\Sigma\, x_1x_2}{n^2(n-1)} - 2\,\frac{\Sigma\, x_1x_2\,\Sigma\, x_1^2}{n^2(n-1)} + 4\,\frac{(\Sigma\, x_1x_2)^2}{n^2(n-1)^2}, \qquad \ldots \quad (4.2)$$

and (22) can be split up in like manner into

$$\begin{array}{cc} 2 & 2 \\ \hline 2 & 2 \end{array} \qquad \begin{array}{cc} 2 & 1 \\ & 1 \\ \hline 2 & 2 \end{array} \qquad \begin{array}{cc} 1 & 2 \\ 1 & \\ \hline 2 & 2 \end{array} \qquad \begin{array}{cc} 1 & 1 \\ 1 & 1 \\ \hline 2 & 2 \end{array}. \qquad \ldots \quad (4.3)$$

The next step is to replace the products of monomial symmetric functions by single monomial functions. Thus

$$(\Sigma\, x_1^2)^2 = \Sigma\, x_1^4 + 2\,\Sigma\, x_1^2x_2^2, \qquad \ldots \quad (4.4\mathrm{a})$$

$$\Sigma\, x_1^2\,\Sigma\, x_1x_2 = \Sigma\, x_1^3x_2 + \Sigma\, x_1^2x_2x_3 = \Sigma\, x_1x_2\,\Sigma\, x_1^2, \qquad \ldots \quad (4.4\mathrm{b},\ 4\mathrm{c})$$

$$(\Sigma\, x_1x_2)^2 = \Sigma\, x_1^2x_2^2 + 2\,\Sigma\, x_1^2x_2x_3 + 6\,\Sigma\, x_1x_2x_3x_4\,. \qquad \ldots \quad (4.4\mathrm{d})$$

These correspond to the bipartitions, or arrays,

$$\underset{(\Sigma\, x_1^4)}{\begin{array}{cc|c} 2 & 2 & 4 \\ \hline 2 & 2 & 4 \end{array}} \qquad \underbrace{\begin{array}{cc|c} 2 & & 2 \\ & 2 & 2 \\ \hline 2 & 2 & 4 \end{array} \qquad \begin{array}{cc|c} & 2 & 2 \\ 2 & & 2 \\ \hline 2 & 2 & 4 \end{array}}_{(\Sigma\, x_1^2 x_2^2)} \qquad \cdots \quad (4.5a)$$

$$\underset{(\Sigma\, x_1^3 x_2)}{\begin{array}{cc|c} 2 & 1 & 3 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \qquad \underset{(\Sigma\, x_1^2 x_2 x_3)}{\begin{array}{cc|c} 2 & & 2 \\ & 1 & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \qquad \cdots \quad (4.5b)$$

$$\underset{(\Sigma\, x_1^3 x_2)}{\begin{array}{cc|c} 1 & 2 & 3 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array}} \qquad \underset{(\Sigma\, x_1^2 x_2 x_3)}{\begin{array}{cc|c} & 2 & 2 \\ 1 & & 1 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array}} \qquad \cdots \quad (4.5c)$$

$$\underset{(\Sigma x_1^2 x_2^2)}{\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & 1 & 2 \\ \hline 2 & 2 & 4 \end{array}} \quad \underbrace{\begin{array}{cc|c} 1 & 1 & 2 \\ 1 & & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} 1 & 1 & 2 \\ & 1 & 1 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array}}_{(\Sigma x_1^2 x_2 x_3)} \quad \underbrace{\begin{array}{cc|c} 1 & & 1 \\ 1 & & 1 \\ & 1 & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} 1 & & 1 \\ & 1 & 1 \\ 1 & & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} 1 & & 1 \\ & 1 & 1 \\ & 1 & 1 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} & 1 & 1 \\ 1 & & 1 \\ 1 & & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} & 1 & 1 \\ 1 & & 1 \\ & 1 & 1 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array} \quad \begin{array}{cc|c} & 1 & 1 \\ & 1 & 1 \\ 1 & & 1 \\ 1 & & 1 \\ \hline 2 & 2 & 4 \end{array}}_{(\Sigma x_1 x_2 x_3 x_4)} \ \cdots \quad (4.5d)$$

where the partition corresponding to the final symmetric function is written as a column of totals, in a fixed order, and the coefficient of this function in (4.4) is simply the number of different ways of arranging the columns to yield the final partition. This is perfectly general : *the coefficient of* $\Sigma\, x_1^{t_1} \ldots x_\gamma^{t_\gamma}$ *in* $(\Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha})_1 \ldots (\Sigma x_1^{a_1} \ldots x_\alpha^{a_\alpha})_\rho$ *is equal to the number of arrays in which the* ρ *columns, which are written in a fixed order, are permutations of* $(a_1 \ldots a_\alpha)_1, \ldots, (a_1 \ldots a_\alpha)_\rho$ *respectively, and the row totals are* $t_1, \ldots, t_\gamma$ *written in a fixed order*. The truth of this rule can easily be seen if one considers the correspondence between the algebraic process of evaluating the product of monomial symmetric functions, and the process of writing down arrays subject to these conditions.[1] Although a formal proof may be constructed it is not very illuminating.

---

[1] If the reader will evaluate few products like $\Sigma\, x_1^2 x_2 x_3\ \Sigma\, x_1^2 x_2$ ($= \Sigma\, x_1^4 x_2^2 x_3 + 3\, \Sigma\, x_1^4 x_2 x_3 x_4 + 2\, \Sigma\, x_1^3 x_2^3 x_3 + 2\, \Sigma\, x_1^3 x_2^2 x_3^2 + 3\, \Sigma\, x_1^3 x_2^2 x_3 x_4 + 6\, \Sigma\, x_1^2 x_2^2 x_3^2 x_4 + 6\, \Sigma x_1^2 x_2^2 x_3 x_4 x_5$) he will probably find himself using just this sort of diagram, but with rows and columns interchanged. But here we stick to the convention established by Fisher.

Collecting together the terms given by (4.2) and (4.4) we have

$$k_2^2 = \frac{1}{n^2}(\mathit{1}\,\Sigma\, x_1^4 + \mathit{2}\,\Sigma\, x_1^2 x_2^2) - \frac{2}{n^2(n-1)}\,(\mathit{2}\,\Sigma\, x_1^3\, x_2 + \mathit{2}\,\Sigma\, x_1^2 x_2 x_3) \qquad \dots \quad (4.6)$$

$$+\frac{4}{n^2(n-1)^2}\,(\mathit{1}\,\Sigma\, x_1^2 x_2^2 + \mathit{2}\,\Sigma\, x_1^2 x_2 x_3 + \mathit{6}\,\Sigma\, x_1 x_2 x_3 x_4)$$

$$= \mathit{1}\frac{1}{n^2}\;\Sigma\, x_1^4 + \mathit{2}\;\frac{-2}{n^2(n-1)}\;\Sigma\, x_1^3 x_2 + \left(\mathit{2}\frac{1}{n^2} + \mathit{1}\;\frac{4}{n^2(n-1)^2}\right)\Sigma\, x_1^2 x_2^2$$

$$+\left(\mathit{2}\;\frac{-2}{n^2(n-1)} + \mathit{2}\;\frac{4}{n^2(n-1)^2}\right)\Sigma x_1^2 x_2 x_3 + \mathit{6}\;\frac{4}{n^2(n-1)^2}\;\Sigma\, x_1 x_2 x_3 x_4,$$

the seven detailed terms in (4.6) corresponding to the typical arrays

$$\underset{\mathit{1}}{\begin{array}[b]{cc|c} 2 & 2 & 4 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{2}}{\begin{array}[b]{cc|c} 2 & 1 & 3 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{2}}{\begin{array}[b]{cc|c} 2 & & 2 \\ & 2 & 2 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{1}}{\begin{array}[b]{cc|c} 1 & 1 & 2 \\ 1 & 1 & 2 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{2}}{\begin{array}[b]{cc|c} 2 & & 2 \\ & 1 & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{2}}{\begin{array}[b]{cc|c} 1 & 1 & 2 \\ 1 & & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \quad \underset{\mathit{6}}{\begin{array}[b]{cc|c} 1 & & 1 \\ 1 & & 1 \\ & 1 & 1 \\ & 1 & 1 \\ \hline 2 & 2 & 4 \end{array}} \qquad \dots \quad (4.7)$$

and the italic figures denoting the number of arrays having the same column and row totals, and the same columns apart from permutations within and between columns. (The permutations *between* columns arise through the combination of terms like $\Sigma x_1^2\; \Sigma x_1 x_2$ and $\Sigma x_1 x_2 \Sigma x_1^2$, that is to say of arrays like (4.5*b*) and (4.5*c*) i.e. (45c)). The general result is, using (4.1).

Theorem 4.1 : *The coefficient of* $\Sigma\, x_1^{t_1} \dots x_\tau^{t_\tau}$ *in* $z_{r_1} \dots z_{r_\rho}$ $(\,\Sigma\, r_h = w = \Sigma\, t_j)$ *is the sum, over all the different bipartitions, (a), of the two partitions* $(r_1 \,\dots\, r_\rho)$ *and* $(t_1 \dots t_\tau)$*, of the quantities*

$$\prod_{h=1}^{\rho} \left[\binom{r}{a_1 \dots a_\alpha} B_{(z)}(a_1 \dots a_\alpha) \right]_h . \qquad \dots \quad (4.8)$$

Since

$$E\;\Sigma\,(x_1 \dots x_{\tau_1})^{t_1} \dots (x_{\tau-\tau_T+1} \cdots x_\tau)^{t_T} = \frac{n^{(\tau)}}{\tau_1\,! \dots \tau_T!}\,\mu'^{\tau_1}_{t_1} \;\dots\, \mu'^{\tau_T}_{t_T}, \qquad \dots \quad (4.9)$$

or, in extended notation,

$$E\;\Sigma\, x_1^{t_1} \dots x_\tau^{t_\tau} = \frac{n^{(\tau)}}{\tau_1! \,\dots\, \tau_T\,!}\;\mu'_{t_1} \cdots \mu'_{t_\tau}, \qquad \dots \quad (4.10)$$

we have the following theorem.

Theorem 4.2 : *The coefficient of* $\mu'_{t_1} \dots \mu'_{t_\tau}$ *in* $\mu'_{(z)}(r_1 \dots r_\rho)$ *is the sum, over all the different bipartitions,* (*a*), *of the two partiions* $(r_1 \dots r_\rho)$ *and* $(t_1 \dots t_\tau)$, *of the quantities*

$$\prod_h \left[\binom{r}{a_1 \dots a_\alpha} B_{(z)}(a_1 \dots a_\alpha) \right]_h \cdot \frac{n^{(\tau)}}{\tau_1! \dots \tau_T!} \qquad \dots \quad (4.11)$$

$$= \frac{1}{\tau_1! \dots \tau_T!} \prod_h \binom{r}{a_1 \dots a_\alpha}_h \times n^{(\tau)} \prod_h B_{(z)}(a_1 \dots a_\alpha)_h$$

= (*numerical coefficient*) (μ'-*type pattern function*).

This theorem completes our proof of the fundamental rules for the univariate case of the $\mu' z \mu'$ problem.

Continuing with our example, (4.6) gives

$$\mu'_{(k)}(2^2) = 1 \frac{1}{n^2} n\mu'_4 + 2 \frac{-2}{n^2(n-1)} n(n-1)\mu'_3\mu'_2 + \left(2 \frac{1}{n^2} + 1 \frac{4}{n^2(n-1)^2}\right) \frac{n(n-1)}{2} \mu'^2_2$$

$$+ \left(2 \frac{-2}{n^2(n-1)} + 2 \frac{4}{n^2(n-1)^2}\right) \frac{n(n-1)(n-2)}{2} \mu'_2 \mu'^2_1 \qquad \dots \quad (4.12)$$

$$+ 6 \frac{4}{n^2(n-1)^2} \frac{n(n-1)(n-2)(n-3)}{24} \mu'^4_2$$

$$= \frac{1}{n} \mu'_4 - \frac{4}{n} \mu'_3\mu'_1 + \frac{n^2-2n+3}{n(n-1)} \mu'^2_2 - \frac{2(n-2)(n-3)}{n(n-1)} \mu'_2\mu'^2_1 + \frac{(n-2)(n-3)}{n(n-1)} \mu'^4_1.$$

If we take $\mu'_1 = 0$, this reduces to

$$\mu'_{(k)}(2^2) = \frac{1}{n} \mu_4 + \frac{n^2-2n+3}{n(n-1)} \mu^2_2. \qquad \dots \quad (4.13)$$

The proof of the fundamental rules for the multivariate case is on just the same lines. The multilinear case is particularly interesting, and particularly simple. We shall follow Kaplan in replacing multipartite numbers consisting of strings of zeros and units by numbers giving the positions of the units. For example, the partition ((1010)(0001)) of (1011) will be spoken of as the partition [13, 4] of [134], and the partitions corresponding to the typical terms on the right-hand sides of (3.21) and (3.27) are [*ijkl*], [*ijk*,*l*], [*ij*, *kl*].[*ij*, *k*, *l*], [*i*,*j*, *k*, *l*]. Now when multiplied together the monomial symmetric functions $\Sigma\, x_1^{i_1} \dots x_1^{i_{a_1}} \dots x_\alpha^{i_{r-a_\alpha+1}} \dots x_\alpha^{i_r}$ behave just like the univariate ones. Thus, to evaluate the product $\Sigma\, x_1^i\, x_1^j\, \Sigma\, x_1^k\, x_2^l$ we write down (in index notation) the bipartitions

$$\begin{array}{cc|c} ij & k & ijk \\ & l & l \\ \hline ij & kl & ijkl \end{array} \qquad \begin{array}{cc|c} ij & l & ijl \\ & k & k \\ \hline ij & kl & ijkl \end{array} \qquad \begin{array}{cc|c} ij & & ij \\ & k & k \\ & l & l \\ \hline ij & kl & ijkl \end{array} \qquad \ldots \quad (4.14)$$

which summarize the result

$$\Sigma\, x_1^i x_1^j\, \Sigma\, x_1^k x_2^l = \Sigma\, x_1^i x_1^j x_1^k x_2^l + \Sigma\, x_1^i x_1^j x_1^l x_2^k + \Sigma\, x_1^i x_1^j x_2^k x_3. \qquad \ldots \quad (4.15)$$

We notice that in the multilinear case there is always a unique bipartition corresponding to given column and row "totals". Moreover the multinomial coefficients present in (3.28) are absent in (3.31), and (4.10) is replaced by

$$E\, \Sigma\, x_1^{j_1} \ldots x_1^{j_{t_1}} \ldots x_\tau^{j_{w-t_\tau+1}} \ldots x_\tau^{j_w}$$

$$= n^{(\tau)}\, \mu'^{j_1 \cdots j_{t_1}} \ldots \mu'^{j_{w-t_\tau+1} \cdots j_w}. \qquad \ldots \quad (4.16)$$

Thus all the numerical coefficients are unity, and Theorem 4.2 for the multilinear case becomes as follows.

Theorem 4.3 : *The coefficient of* $\mu'^{\,j_1 \ldots j_{t_1}} \ldots \mu'^{\,j_{w-t_\tau+1} \cdots j_w}$ *in* $\mu'_{(z)}\, [i_1 \ldots i_{r_1}, \ldots, i_{w-r_\rho+1} \ldots i_w] \equiv E\, z^{i_1 \cdots i_{r_1}} \ldots z^{i_{w-r_\rho+1} \cdots i_w}$ *(*$(j_1 \ldots j_w)$ *a permutation of* $(i_1 \ldots i_w)$*) is given in terms of the unique bipartition of the two partitions* $[i_1 \ldots i_{r_1}, \ldots, i_{w-r_\rho+1} \ldots i_w]$, $[j_1 \ldots j_{t_1}, \ldots, j_{w-t_\tau+1} \ldots j_w]$ *by the* μ'*-type pattern function*

$$n^{(\tau)} \prod_h B_{(z)}\, (a_1 \ldots a_\alpha)_h, \qquad \ldots \quad (4.17)$$

where of course $(a_1 \ldots a_\alpha)_h$ *still denotes the number of indices in the entries of the* h*-th column.*

For example, let us find $\mu'_{(k)}\, [ij, kl]$. Typical bipartitions (compare (4.7)) are

$$\begin{array}{c} \begin{array}{cc|c} ij & kl & ijkl \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^j x_1^k x_1^l) \end{array} \quad \begin{array}{c} \begin{array}{cc|c} ij & k & ijk \\ & l & l \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^j x_1^k x_2^l) \end{array} \quad \begin{array}{c} \begin{array}{cc|c} ij & & ij \\ & kl & kl \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^j x_2^k x_2^l) \end{array} \quad \begin{array}{c} \begin{array}{cc|c} i & k & ik \\ j & l & jl \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^k x_2^j x_2^l) \end{array} \quad \ldots \quad (4.18)$$

$$\begin{array}{c} \begin{array}{cc|c} ij & & ij \\ & k & k \\ & l & l \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^j x_2^k x_3^l) \end{array} \quad \begin{array}{c} \begin{array}{cc|c} i & k & ik \\ j & & j \\ & l & l \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_1^k x_2^j x_3^l) \end{array} \quad \begin{array}{c} \begin{array}{cc|c} i & & i \\ j & & j \\ & k & k \\ & l & l \\ \hline ij & kl & ijkl \end{array} \\ (\Sigma\, x_1^i x_2^j x_3^k x_4^l) \end{array}.$$

J-393
7.9.70

Thus (3.25) yields

$$k^{ij}\,k^{kl} = \Big(\frac{\Sigma\, x_1^i x_1^j}{n} - \frac{\Sigma\, x_1^i x_2^j}{n^{(2)}}\Big)\Big(\frac{\Sigma\, x_1^k x_1^l}{n} - \frac{\Sigma\, x_1^k x_2^l}{n^{(2)}}\Big)$$

$$= \frac{1}{n^2}\,\Sigma\, x_1^i x_1^j x_1^k x_1^l - \frac{1}{n^2(n-1)}\,\overset{4}{\Sigma}\,\Sigma\, x_1^i x_1^j x_1^k x_2^l +$$

$$+\frac{1}{n^2}\,\Sigma\, x_1^i x_1^j x_2^k x_2^l + \frac{1}{n^2(n-1)^2}\,(\Sigma\, x_1^i x_1^k x_2^j x_2^l + \Sigma\, x_1^i x_1^l x_2^j x_2^k) -$$

$$-\frac{1}{n^2(n-1)}\,\overset{2}{\Sigma}\,\Sigma\, x_1^i x_1^j x_2^k x_3^l + \frac{1}{n^2(n-1)^2}\,\overset{4}{\Sigma}\,\Sigma\, x_1^i x_1^k x_2^j x_3^l +$$

$$+\frac{1}{n^2(n-1)^2}\,\Sigma\, x_1^i x_2^j x_3^k x_4^l. \qquad \ldots \quad (4.19)$$

Thus, as given by Theorem 4.3,

$$\mu'_{(k)}[ij,\, kl] = \frac{1}{n}\mu'^{ijkl} - \frac{1}{n}\,\overset{4}{\Sigma}\,\mu'^{ijk}\,\mu'^{l} + \frac{n-1}{n}\,\mu'^{ij}\,\mu'^{kl} +$$

$$+\frac{1}{n(n-1)}\,(\mu'^{ik}\,\mu'^{jl} + \mu'^{il}\,\mu'^{jk}) - \frac{n-2}{n}\,\overset{2}{\Sigma}\,\mu'^{ij}\,\mu'^{k}\,\mu'^{l} +$$

$$+\frac{n-2}{n(n-1)}\,\overset{4}{\Sigma}\,\mu'^{ik}\,\mu'^{j}\,\mu'^{l} + \frac{(n-2)(n-3)}{n(n-1)}\,\mu'^{i}\,\mu'^{j}\,\mu'^{k}\,\mu'^{l}. \qquad \ldots \quad (4.20)$$

On putting $i = j = k = l$ this reduces to (4.12), and on writing $i = j = k = 1, l = 2$, to

$$\mu'_{(k)}((20)(11)) = \frac{1}{n}\mu'_{31} - \frac{1}{n}\,(\mu'_{30}\,\mu'_{01} + 3\mu'_{21}\,\mu'_{10}) + \frac{n^2-2n+3}{n(n-1)}\,\mu'_{20}\,\mu'_{11} -$$

$$-\frac{(n-2)(n-3)}{n(n-1)}\,(\mu'_{20}\,\mu'_{10}\,\mu'_{01} + \mu'_{11}\,\mu'^{2}_{10}) + \frac{(n-2)(n-3)}{n(n-1)}\,\mu'^{3}_{10}\,\mu'_{01}. \qquad \ldots \quad (4.21)$$

All the proofs given in later sections will be for the multilinear case, and the "numerical coefficients" will all be unity. The numerical coefficients in the univariate and less general multivariate cases arise through the coalescence of terms when we equate some or all of the variates. Since the evaluation of the coefficient of a μ'- or κ-product in a moment or cumulant of z-statistics depends, in each case, on writing down the arrays which have as their column totals the partition appropriate to the moment or cumulant and as their row totals that appropriate to the μ'- or κ-product, the number of coalescing terms will be the same in all three problems ($\mu' z \mu'$, $\mu' z \kappa$, $\kappa z \kappa$). Thus it will be sufficient to evaluate these numerical coefficients once and for all. This we have already done for the univariate case, and, as indicated earlier, there is no essential difficulty in extending the work to the multivariate (but not multilinear) case.

5. PROOF OF THE FUNDAMENTAL RULES FOR THE $\mu' z \kappa$ PROBLEM

The relevant theorems for the univariate and multilinear cases are as follows.

Theorem 5.1 : *The coefficient of* $\kappa_{t_1} \dots \kappa_{t_\tau}$ *in* $\mu'_{(z)}(r_1 \dots r_\rho)$ *is the sum, over all the different bipartitions,* (a), *of the two partitions* $(r_1 \dots r_\rho)$ *and* $(t_1 \dots t_\tau)$, *of the quantities* (numerical coefficient) (κ-type pattern function), *where the numerical coefficient is given by* (4.11). The κ-type pattern function is the sum, over all the separations of the rows of the bipartition (a), of the quantities

$$n^{(\chi)} \prod_h B_{(z)}(b_1 \dots b_\beta)_h, \qquad \dots \quad (5.1)$$

where (b) is the specification of such a separation, and $(q_1 \dots q_\chi)$ is the specification of the corresponding separation of the row totals.

Theorem 5.2 : *The coefficient of* $\kappa^{j_1 \dots j_{t_1}} \dots \kappa^{j_{w-t_\tau+1} \dots j_w}$ *in* $\mu'_{(z)}[i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \dots i_w]$ *is given in terms of the unique bipartition of the two partitions* $[i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \dots i_w]$ *and* $[j_1 \dots j_{t_1}, \dots, j_{w-t_\tau+1} \dots j_w]$ *by the* κ*-type pattern function defined in Theorem* 5.1. (The a's now refer to the number of indices in the cells of the actual partition.)

There is also the intermediate "multivariate" case, which we shall not enunciate explicitly.

We confine ourselves to proving Theorem 5.2; Theorem 5.1 will then follow on the lines indicated at the end of section 4.

According to Theorem 4.3, the coefficient of $\mu'^{l_1 \dots l_{q_1}} \dots \mu'^{l_{w-q_\chi+1} \dots l_w}$ in $\mu'_{(z)}[i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \dots i_w]$ is given by (5.1), where the b's refer to the unique bipartition of $[i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \dots i_w]$ and $[l_1 \dots l_{q_1}, \dots, l_{w-q_\chi+1}, \dots l_w]$. Now when $\mu'^{l_1 \dots l_{q_1}} \dots \mu'^{l_{w-q_\chi+1} \dots l_w}$ is expressed in terms of cumulants, we see from (3.24) that we get the sum of terms like $\kappa^{j_1 \dots j_{t_1}} \dots \kappa^{j_{w-t_\tau+1} \dots j_w}$, where each group of j's is a part (or the whole) of one group of l's. In other words $[j_1 \dots j_{t_1}, \dots, j_{w-t_\tau+1} \dots j_w]$ is a *sub-partition* of $[l_1 \dots l_{q_1}, \dots, l_{w-q_\chi+1} \dots l_w]$. For example,

$$\begin{aligned} \mu'^{ijk} \mu'^{lm} &= (\kappa^{ijk} + \kappa^{ij}\kappa^{k} + \kappa^{ik}\kappa^{j} + \kappa^{jk}\kappa^{i} + \kappa^{i}\kappa^{j}\kappa^{k})(\kappa^{lm} + \kappa^{l}\kappa^{m}) \\ &= \kappa^{ijk} + \kappa^{lm} + \kappa^{ijk}\kappa^{l}\kappa^{m} + \dots + \kappa^{i}\kappa^{j}\kappa^{k}\kappa^{l}\kappa^{m}. \qquad \dots \quad (5.2) \end{aligned}$$

But this is only another way of saying that a term in $\kappa^{j_1 \dots j_{t_1}} \dots \kappa^{j_{w-t_\tau+1} \dots j_w}$ arises from each term in $\mu'^{l_1 \dots l_{q_1}} \dots \mu'^{l_{w-q_\chi+1} \dots l_w}$ in which the l-groups arise by amalgamation of the j-groups; or, in other words, the relevant partitions $[l_1 \dots l_{q_1}, \dots, l_{w-q_\chi+1} \dots l_w]$ are those which are the specifications of separations of $[j_1 \dots j_{t_1}, \dots, j_{w-t_\tau+1} \dots j_w]$. For example, the term in $\kappa^{ijk}\kappa^{l}\kappa^{m}$ arises from the term in $\mu'^{ijk}\ \mu'^{lm}$ (by (5.2)) and in addition from those in $\mu'^{ijk}\,\mu'^{l}\,\mu'^{m}$, $\mu'^{ijkl}\,\mu'^{m}$, $\mu'^{ijkm}\,\mu'^{l}$ and μ'^{ijklm}, and from no others. Since each μ'-product has its own coefficient (5.1), and all the coefficients in the relation (3.24) are unity, the theorem follows.

As an example, let us consider the coefficient of $\kappa^{ik}\ \kappa^{jl}$ in $\mu'_{(k)}\,[ij, kl]$. In (4.20) the only terms giving rise to $\kappa^{ik}\ \kappa^{jl}$ are those in $\mu'^{ik}\ \mu'^{jl}$ and μ'^{ijkl} (corresponding to the two specifications, $[ik, jl]$ and $[ijkl]$, of the partition $[ik, jl]$). Thus the required coefficient is $1/n(n-1)+1/n = 1/(n-1)$. If we care to start right from the beginning and use Theorem 5.2 in connection with the arrays (4.18), we obtain

$$\mu'_{(k)}\,[ij, kl] = n\,\frac{1}{n}\,\frac{1}{n}\,\kappa^{ijkl}+\left(n^{(2)}\,\frac{1}{n}\,\frac{1}{n}+n\,\frac{1}{n}\,\frac{1}{n}\right)\kappa^{ij}\,\kappa^{kl}+$$

$$+\left(n^{(2)}\,\frac{-1!}{n^{(2)}}\,\frac{-1!}{n^{(2)}}+n\,\frac{1}{n}\,\frac{1}{n}\right)(\kappa^{ik}\,\kappa^{jl}+\kappa^{il}\,\kappa^{jk})$$

$$= \frac{\kappa^{ijkl}}{n}+\kappa^{ij}\,\kappa^{kl}+\frac{\kappa^{ik}\kappa^{jl}+\kappa^{il}\kappa^{jk}}{n-1}. \qquad \dots \quad (5.3)$$

(Terms containing first order κ's are absent by Rule 2.) If $i = j = k = l$ this reduces to the univariate result

$$\mu'_{(k)}\,(2^2) = \frac{\kappa_4}{n}+\frac{n+1}{n-1}\,\kappa_2^2, \qquad \dots \quad (5.4)$$

also obtainable directly by using the arrays (4.7) and Theorem 5.1. Similarly, putting $i = k = 1$, $j = l = 2$, we find

$$\mu'_{(k)}\,((11)^2) = \frac{\kappa_{22}}{n}+\kappa_{11}^2+\frac{\kappa_{20}\kappa_{02}+\kappa_{11}^2}{n-1}. \qquad \dots \quad (5.5)$$

6. PROOF OF THE FUNDAMENTAL RULES FOR THE $\kappa z \kappa$ PROBLEM

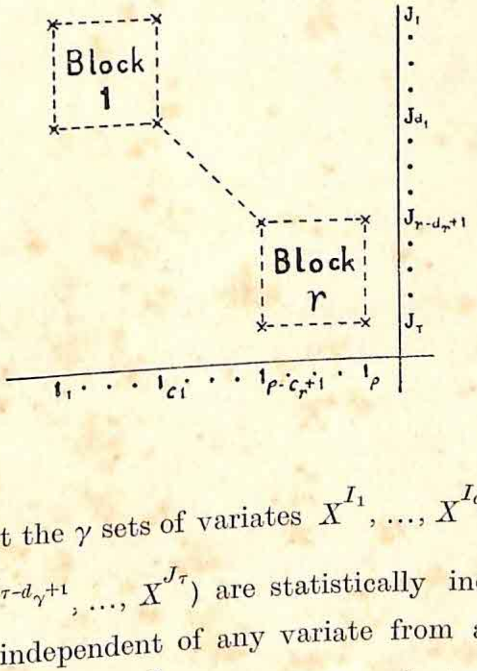
Theorem 6.1 : *The coefficient of* $\kappa_{t_1} \dots \kappa_{t_\tau}$ *in* $\kappa_{(z)}\,(r_1 \dots r_\rho)$ *(univariate case) or the coefficient of* $\kappa^{j_1 \cdot\cdot j_{t_1}} \dots \kappa^{j_{w-t_\tau+1} \cdot\cdot j_w}$ *in* $\kappa_{(z)}\,[i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \dots i_w]$ *(multilinear case) is given by Theorem* 5.1 *or Theorem* 5.2, *as the case may be, except that bipartitions which split into two or more complete blocks, each confined to separate rows and columns, are to be neglected.*

We make this theorem depend on the following lemma, which also justifies Rule 3 ($\mu' z \kappa$).

Lemma 6.1 : *The* κ*-type pattern function of any such bipartition is equal to the product of the* κ*-type pattern functions of its separate blocks.*

Again we may confine ourselves to the multilinear case. We remark that in order to prove Theorem 6.1 (and hence the rules for the $\kappa z \kappa$ problem) it is not sufficient merely to demonstrate the absence of terms arising from dissectable partitions (a proof of the latter was indicated by Kaplan (1952)); it is also necessary to show that the remaining terms are the same as for the $\mu' z \kappa$ problem, and are given by Theorem 5.1 or 5.2.

First we prove the lemma, and begin by introducing an abbreviation. We shall denote the sets of indices $i_1 \dots i_{r_1}, \dots, i_{w-r_\rho+1} \cdots i_w$ by $I_1, \dots, I_\rho$ and the sets $j_1 \cdots j_{t_1}, \dots, j_{w-t_\tau+1} \cdots j_w$ by $J_1, \dots, J_\tau$. We may now speak of the coefficient of $\kappa^{J_1} \dots \kappa^{J_\tau}$ in $\kappa_{(z)}[I_1, \dots, I_\rho]$, and so on. We shall also write $X^{I_1}; \dots; X^{I_\rho}$ for the sets of variates $x^{i_1}, \dots, x^{i_{r_1}}; \dots; x^{i_{w-r_\rho+1}}, \dots, x^{i_w}$ and $X^{J_1}; \dots; X^{J_\tau}$ for the sets $x^{j_1}, \dots, x^{j_{t_1}}; \dots; x^{j_{w-t_\tau+1}}, \dots, x^{j_w}$. By a rearrangement of columns and rows, which does not affect the pattern function, we may always ensure that the γ blocks of a dissectable partition run diagonally, as in the Figure; that is to say the 1st block is confined to the first c_1 columns and d_1 rows, ..., the γ-th block to the last c_γ columns and d_γ rows.



Now suppose that the γ sets of variates $X^{I_1}, \dots, X^{I_{c_1}}; \dots; X^{I_{\rho-c_\gamma+1}}, \dots, X^{I_\rho}$ $(\equiv X^{J_1}, \dots, X^{J_{d_1}}; \dots; X^{J_{\tau-d_\gamma+1}}, \dots, X^{J_\tau})$ are statistically independent (so that any variate from one set is independent of any variate from another set). Then the products $z^{I_1} \dots z^{I_{c_1}}, \dots, z^{I_{\rho-c_\gamma+1}} \dots z^{I_\rho}$ are also independent, so that

$$\mu'_{(z)}[I_1, \dots, I_\rho] = \mu'_{(z)}[I_1, \dots, I_{c_1}] \dots \mu'_{(z)}[I_{\rho-c_\gamma+1}, \dots, I_\rho]. \qquad \dots (6.1)$$

Now when the moments on the right-hand side of (6.1) are expressed in terms of population cumulants, κ^J, each group of indices J must obviously arise from *one only* of the γ groups $I_1 \ldots I_{c_1}, \ldots, I_{\rho-c_\gamma+1} \ldots I_\rho$ (see Figure 6.1). The same applies to the left-hand side, this time because the assumed independence of the γ groups of variates makes any κ^J spanning more than one of the γ groups vanish.[1] But the remaining κ^J may still be chosen arbitrarily, so that we are entitled to equate coefficients on either side of (6.1). Thus the coefficient of $\kappa^{J_1} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_1, \ldots, I_\rho]$ is the product of the coefficients of $\kappa^{J_1} \ldots \kappa^{J_{d_1}}$ in $\mu'_{(z)}[I_1, \ldots, I_{c_1}], \ldots, \kappa^{J_{\tau-d_\gamma+1}} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_{\rho-c_\gamma+1}, \ldots, I_\rho]$. Since the numerical coefficients are unity (see Theorem 5.2) this proves the lemma.

Let us now consider Theorem 6.1 itself. Equation (3.24), written for moments and cumulants of z-statistics, is

$$\mu'_{(z)}[I_1, \ldots, I_\rho] = \sum_{(c)}\ \sum_{p\rho} \kappa_{(z)}[I_1, \ldots, I_{c_1}] \ldots \kappa_{(z)}[I_{\rho-c_\gamma+1} \ldots, I_\rho] \qquad \ldots \quad (6.2)$$

$$= \kappa_{(z)}[I_1, \ldots, I_\rho] + \kappa_{(z)}[I_1, \ldots, I_{\rho-1}]\kappa_{(z)}[I_\rho] + \ldots + \kappa_{(z)}[I_1] \ldots \kappa_{(z)}[I_\rho],$$

where the first sum is over all partitions $(c_1 \ldots c_\gamma)$ of ρ and the second over all distinct terms obtained by permuting $I_1, \ldots, I_\rho$. The only term on the right of (6.2) involving only a single $\kappa_{(z)}$ is $\kappa_{(z)}[I_1, \ldots, I_\rho]$ itself; the others are products of $\kappa_{(z)}$'s which are individually of lower order. Consequently if by assuming the theorem is true for these lower order $\kappa_{(z)}$'s we can prove that it is true for $\kappa_{(z)}[I_1, \ldots, I_\rho]$ itself, we have the basis for a proof by induction. For the theorem is certainly true for $\kappa_{(z)}[I_1], \ldots, \kappa_{(z)}[I_\rho]$, since no bipartition corresponding to $\kappa_{(z)}[I]$ can possibly split into two or more blocks, while $\kappa_{(z)}[I]$ is the same as $\mu'_{(z)}[I]$, to which Theorem 5.2 applies.

To show that the theorem is true for $\kappa_{(z)}[I_1 \ldots, I_\rho]$ if it is true for the $\kappa_{(z)}$'s of lower order, it is sufficient to prove, for any partitions $[I_1, \ldots, I_\rho]$, $[J_1, \ldots, J_\tau]$ determining a bipartition which falls into $\gamma(\geqslant 2)$ blocks as in the Figure that the coefficient of $\kappa^{J_1} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_1, \ldots, I_\rho]$ is equal to the product of the coefficients of $\kappa^{J_1} \ldots \kappa^{J_{d_1}}$ in $\kappa_{(z)}[I_1, \ldots, I_{c_1}], \ldots, \kappa^{j_{\tau-d_\gamma+1}} \ldots \kappa^{J_\tau}$ in $\kappa_{(z)}[I_{\rho-c_\gamma+1}, \cdots, I_\rho]$. For then it follows, by (6.2), that if the bipartition consists of one block only, the coefficients of $\kappa^{J_1} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_1, \ldots, I_\rho]$ and $\kappa_{(z)}[I_1, \ldots, I_\rho]$ are equal ; in other words, $\kappa_{(z)}[I_1, \cdots, I_\rho]$ is given by those terms of $\mu'_{(z)}[I_1, \ldots, I_\rho]$ which correspond to nondis- sectable bipartitions, which is what we require to prove. Finally, since we are assum- ing the theorem for $\kappa_{(z)}$'s of lower order, and individual blocks are no longer dissec- table, we only have to show that the coefficient of $\kappa^{J_1} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_1, \ldots, I_\rho]$ is equal to the product of the coefficients of $\kappa^{J_1} \ldots \kappa^{J_{d_1}}$ in $\mu'_{(z)}[I_1, \ldots, I_{c_1}], \ldots, \kappa^{J_{\tau-d_\gamma+1}} \ldots \kappa^{J_\tau}$ in $\mu'_{(z)}[I_{\rho-c_\gamma+1}, \ldots, I_\rho]$, and this is so by virtue of the lemma.

---

[1] It is at this point that the "proof" for the corresponding, untrue, result for μ'-type pattern functions would break down. For example if x^i and x^j are independent we have $\mu'_{(k)}[i, j] = \mu'_{(k)}[i]\,\mu'_{(k)}[j]$ or $\mu'^{ij}/n + (n-1)\mu'^{i}\mu'^{j}/n = \mu'^{i}\mu'^{j}$, but we are not thereby entitled to equate the coefficients of $\mu'^{i}\mu'^{j}$, for $\mu'^{ij} = \mu'^{i}\mu'^{j}$ when x^i and x^j are independent.

7. Proofs of the auxiliary rules

Theorem 7.1 : (*Rule* 2, *multilinear form.*) *If the z-statistics are seminvariant, then except for moments involving first-order z-statistics* (z^i) *and the single* $\kappa z\kappa$ *result* $\kappa_{(z)}[i] \propto \kappa^i$, *no* $\mu' z\kappa$ *or* $\kappa z\kappa$ *formula involves the first-order population cumulants*, κ^j.

For since the z-statistics concerned (except the z^i) are invariant under a change of origin, so are all the moments of z-statistics not involving a z^i, and all the cumulants of z-statistics except the $\kappa_{(z)}[i]$. Thus all these results are independent of the values of the κ^j.

Rules 7 and 8 (sections 2 and 3) are often enunciated without including the clause, "not forming a block by itself." In this form they are untrue. For example, the κ-type pattern function (for k-statistics) of the pattern of non-zero elements

$$\begin{array}{ccc} \times & \times & \\ \times & \times & \\ & & \times \end{array} \qquad \ldots \quad (7.1)$$

is neither zero, as required by Rule 7, nor $1/n(n-1)$, as required by Rule 8, but $1/(n-1)$. (This is most easily seen by using the product rule for dissectable arrays, Lemma 6.1.) Of coure such an array is to be ignored in the $\kappa k\kappa$ problem, by Rule 3. We shall give a separate proof of Rule 7 (which is a particular case of Rule 9) because our proof of Rule 9 involves the use of Rule 7.

Theorem 7.2 : (Rule 7). *The* κ*-type pattern function* (*for k-statistics* only) *of any bipartition containing a row consisting of a single non-zero entry, not forming a block by itself, vanishes.*

We assume the bipartition is not dissectable. (If it is, use Lemma 6.1 first). We may make all the non-zero entries correspond to different *single* variates. We shall then have one row total corresponding to a first-order cumulant, κ^j, and at least one column total corresponding to a k-statistic of order greater than one. The corresponding term in the $\kappa k\kappa$ problem vanishes by Rule 2; but the bipartition is not one which has to be neglected according to Rule 3, the numerical coefficient is unity, and *the above bipartition is the only one yielding the term considered.* Therefore its pattern function vanishes.

It will be seen that this is a slight revision of the proof given by Kendall (1940, 1947), the use of the multilinear argument ensuring the truth of the statement italicized above.

Rule 8 is an easy corollary of the fundamental Rule 6, and a formal proof will be omitted.

Theorem 7.3 : (Rule 9). *The* κ*-type pattern function* (*for k-statistics* only) *of a bipartition consisting of two blocks, unconnected except by precisely one column, vanishes.*

In proving the theorem we may restrict ourselves to cases, such as

$$\begin{array}{cccc} \times & \times & & \\ \times & \times & & \\ & & \times & \times \\ & \times & \times & \times \end{array} \quad \text{and} \quad \begin{array}{ccc} \times & & \times \\ \times & \times & \\ \times & & \times \\ \times & \times & \end{array} \qquad \dots \quad (7.2)$$

which cannot be dealt with by the simpler Rule 7. Now the pattern function of *any* bipartition (which has no row containing only a single non-zero entry) may be expanded linearly in terms of the pattern functions of the arrays obtained by omitting one column and amalgamating all, some or none of the rows in which non-zero elements were present in the row omitted. For example, the pattern function of the first array in (7.2) may be expanded linearly in terms of those of

$$\begin{array}{ccc} \times & & \\ & \times & \times \\ \times & \times & \times \end{array} \quad \text{and} \quad \begin{array}{ccc} \times & & \\ \times & & \\ & \times & \times \\ \times & \times & \times \end{array} \qquad \dots \quad (7.3)$$

obtained by omitting the first column, the coefficients being $1/(n-1)$ and $-1/n(n-1)$ respectively. Fisher and Wishart (1932) first derived the formulae for such expansions,[1] and they are also considered by Kendall (1947) in connection with this same Rule 9. The reader is referred to one of these authors for further details. In the case of a bipartition of the type considered in the theorem, if we reduce it by successively omitting columns not belonging to both blocks, using Rule 7 whenever possible, the pattern function is seen to vanish.

Theorem 7.4 : (Rule 10, $\kappa z \kappa$ problem, multilinear form). *If*

$$\kappa_{(z)}[I_1, \dots, I_\rho] = \sum_{(t)pw} \sum C[J_1, \dots, J_\tau]\, \kappa^{J_1} \dots \kappa^{J_\tau} \qquad \dots \quad (7.4)$$

(*where the I, J notation is explained in section 6, the second summation is over all different terms obtained by permuting $j_1, \dots, j_w$, and $C[J_1, \dots, J_\tau]$ denotes the coefficient of $\kappa^{J_1} \dots \kappa^{J_\tau}$) then*

$$\kappa_{(z)}[I_1, \dots, I_\rho, h] = B_{(z)}(1) \sum_{(t)pw} \sum C[J_1, \dots, J\ ] \times$$
$$\times (\kappa^{J_1 h} \kappa^{J_2} \dots \kappa^{J_\tau} + \dots + \kappa^{J_1} \dots \kappa^{J_{\tau-1}} \kappa^{J_\tau h}). \qquad \dots \quad (7.5)$$

---

[1] It should be pointed out, however, that the derivation of Fisher and Wishart's expansions may fail if there are rows containing single entries. This is why we have specified the use of Rule 7 whenever possible.

Theorem 7.5 : (*Rule* 10, *$\mu' z\kappa$ problem, multilinear form*). *If*

$$\mu'_{(z)}\,[I_1, \ldots, I_\rho] = \sum_{(t)\mathrm{p}w} \sum C[J_1, \ldots, J_\tau]\, \kappa^{J_1} \ldots \kappa^{J_\tau}, \qquad \ldots \quad (7.6)$$

then

$$\mu'_{(z)}\,[I_1, \ldots, I_\rho, h] = B_{(z)}(1) \sum_{(t)\mathrm{p}w} \sum C[J_1, \ldots, J_\tau] \times$$

$$\times(\kappa^{J_1h} \ldots \kappa^{J_\tau} + \ldots + \kappa^{J_1} \ldots \kappa^{J_\tau h} + n\kappa^{J_1} \ldots \kappa^{J_\tau}\kappa^{h}). \qquad \ldots \quad (7.7)$$

The proofs are easy corollaries of the fundamental rules, in particular of Rule 3 and of Rule 8, and will be omitted. Further details may be found in Kendall (1947). His proof is for the univariate case, but I think that some rather tiresome details that arise in this case have been glossed over. He states, "... This [the addition of a unit column to an array] multiplies all pattern functions by n^{-1}, leaves numerical coefficients unchanged and increases the suffix of every κ according to the row in which the unit appears." A consideration of the case

$$\kappa_{(k)}(2^2) = \frac{\kappa_4}{n} + \frac{2\kappa_2^2}{n-1}, \quad \kappa_{(k)}\,(2^21) = \frac{\kappa_5}{n^2} + \frac{4\kappa_3\kappa_2}{n(n-1)} \qquad \ldots \quad (7.8)$$

will show that this is not the whole story. The numerical coefficients *do* change, but so do the numbers of arrays of the same type, and these changes always seem to compensate one another. As usual, the multilinear case supplies by far the simplest proof, for all numerical coefficients are unity, while there is always just one array corresponding to any given κ-product in the answer.

I would like to thank Professor M. G. Kendall for his interest.

REFERENCES

CHURCH, A. E. R. (1925): On the moments of the distribution of squared standard deviations for samples of n drawn from an indefinitely large population. *Biometrika*, **17**, 79-83.

CRAIG, C. C. (1928): An application of Thiele's semi-invariants to the sampling problem. *Metron*, **7**, 4, 3–74.

FISHER, R. A. (1930): Moments and product moments of sampling distributions. *Proc. London Math. Soc.*, Series 2, **30**, 199–238.

FISHER, R. A. AND WISHART, J. (1932): The derivation of the pattern formulae of two-way partitions from those of simpler patterns. *Proc. London Math. Soc.*, Series 2, **33**, 195–208.

KAPLAN, E. L. (1952): Tensor notation and the sampling cumulants of k-statistics. *Biometrika*, **39**, 319–323.

KENDALL, M. G. (1940): Some properties of k-statistics. *Ann. Eugen.*, **10**, 106–111.

——— (1940A): Proof of Fisher's rules for ascertaining the sampling semi-invariants of k-statistics. *Ann. Eugen.*, **10**, 215–222.

——— (1940B): The derivation of multivariate sampling formulae from univariate formulae by symbolic operation. *Ann. Eugen*, **10**, 392–402.

——— (1942): On seminvariant statistics. *Ann. Eugen.*, **11**, 300–305.

——— (1947): *The Advanced Theory of Statistics*, **1**, 3rd ed., Charles Griffin and Co., London.

Georgescu, N. St. (1932): Further contributions to the sampling problem. *Biometrika*, **24**, 65–107.

Sukhatme, P. V. (1938): On Fisher's combinatorial methods giving moments and cumulants of the distributions of k-statistics. *Proc. Ind. Stat. Con., Calcutta.*

——— (1944): Moments and product moments of moment-statistics for samples of the finite and infinite populations. *Sankhyā*, **6**, 363–382.

Tchouproff, A. A. (1919): On the mathematical expectation of the moments of frequency distributions. *Biometrika*, **12**, 140–169 and 185–210.

——— (1921): On the mathematical expectation of the moments of frequency distributions. *Biometrika*, **13**, 283–295.

Wishart, J. (1933): A comparison of the semi-invariants of the distributions of moment and semi-invariant estimates in samples from an infinite population. *Biometrika*, **25**, 52–60.

Paper received: June, 1956.

A COMPARISON OF SEROLOGICAL AND SOMATOMETRICAL METHODS USED IN DIFFERENTIATING BETWEEN CERTAIN EAST AFRICAN RACIAL GROUPS, WITH SPECIAL REFERENCE TO D^2 ANALYSIS

By D. A. EAST *and* L. OSCHINSKY

Statistical Laboratory, University of Cambridge

SUMMARY. The statistical technique known as Mahalanobis's Generalised Distance, or D^2 analysis was applied to anthropometric data from various East African tribes. It was found that a high degree of discrimination could be achieved, even between closely-related tribes of the Bantu group, by basing the analysis on eight physical characters. The resulting estimates of the 'distance' separating the various racial groups lent little support to the finding of previous investigations of a serological nature.

1. CERTAIN PHYSICAL CHARACTERISTICS OF NEGRO AND NEGROID RACES OF AFRICA

1.1. The problem of the definition of the East African Bantu physical type, or Bantomorph, has concerned anthropologists during the past half-century. Seligman (1913, 1939), Johnston (1902) and Leys and Joyce (1913), have advanced the opinion that the Baganda and other related Bantomorphic tribes of Uganda and British East Africa represent a mixture between a West African racial type (Nigeromorph) or a Central African type (Congomorph), and the negroid Bahima and Batutsi (Hamitomorphs) of southern Uganda and Ruanda-Urundi. Table 1 and Table 2 summarise certain physical characteristics of negro and negroid races of Africa.

TABLE 1. THE DISTRIBUTION OF NON-METRICAL FEATURES IN SOME NEGRO AND NEGROID RACES OF AFRICA

race	hair texture	hair colour	eye colour	skin colour
(1)	(2)	(3)	(4)	(5)
Hamitomorph	curly to frizzly	black	light brown to dark brown	light brown to dark brown
Nilohamitomorph	frizzly	black	dark brown	dark brown to black
Bantomorph	frizzly	black	light brown to dark brown	light brown to black
Nilotomorph	frizzly	black	dark brown	black
Congomorph	frizzly	black	dark brown	light brown to black
East African Mulattomorph	curly to frizzly	black	light brown to dark brown	light brown to dark brown
Nigeromorph	frizzly	black	dark brown	dark brown to black

TABLE 2. THE DISTRIBUTION OF METRICAL FEATURES IN SOME NEGRO AND NEGROID RACES OF AFRICA

race	*(1)*	*(2)*	*(3)*	*(4)*	*(5)*	*(6)*	*(7)*	*(8)*
Hamitomorph	tall	small	long	long	medium	Dolichocephalic	Leptorrhine to mesorrhine	thick
Nilohamitomorph	tall	small	long	long	high	Dolichocephalic	Meso- to platyrrhine	thick
Bantomorph	medium	medium	medium	medium	medium	Dolicho- to mesocephalic	Platyrrhine	thick
Nilotomorph	tall	small	medium to long	long	high	Dolicho- to mesocephalic	Platyrrhine	very thick
Congomorph	medium to short	large	short	short	low	Meso-to brachycephalic	Platyrrhine	thick
East African Mulattomorph	variable	large	short to medium	short to medium	medium	Meso- to brachycephalic	Mesorrhine	medium
Nigeromorph	medium	medium	—	—	high	Mesocephalic	Platyrrhine	thick

Key: (1) Stature ; (2) Sitting Height relative to Stature ; (3) Total Upper Extremity Length relative to Sitting Trunk Height ; (4) Total Lower Extremity Length relative to Sitting Trunk Height ; (5) Forearm Length relative to Upper Arm Length (6) Head Breadth relative to Head Length ; (7) Nose Breadth relative to Nose Height ; (8) Lip Thickness

2. Results of the anthropometric survey of 1950-51

2.1. During 1950-51, one of the authors, L. Oschinsky, carried out an anthropometric survey of various tribes in the area of East Africa depicted in

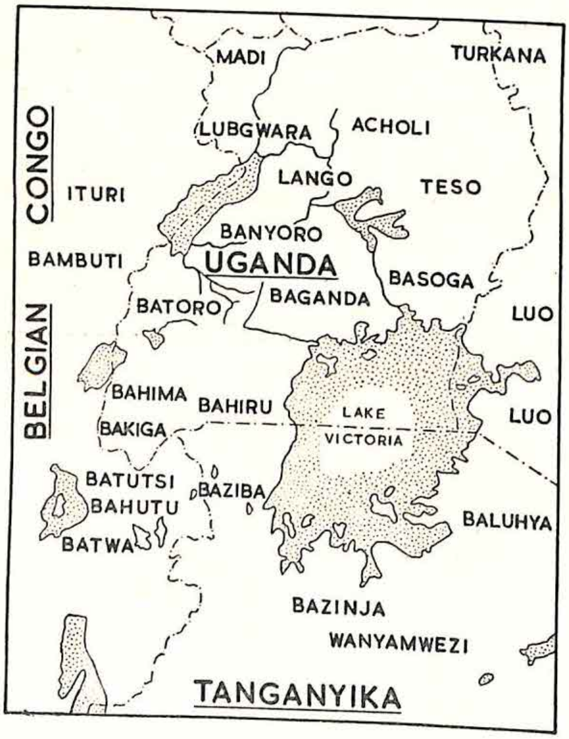


Fig. 1. Distribution of the Uganda, Tanganyika, Kenya, Ruanda-Urundi and Belgian Congo tribes in the survey area.

Figure 1, special attention being paid to the Baganda Bantomorphs and Batutsi Hamitomorphs (Oschinsky, 1954). Preliminary statistical examination of the data, based on *t*-tests, showed that these two groups differed significantly with regard to the majority of the characters measured.

2.2. ABO blood group and sicklaemia data, compiled by other investigators, were considered in connection with the somatometrical and statistical results of the survey. The sicklaemia results were found to be generally in agreement with the somatometry. The ABO blood group data, however, despite the claims of certain proponents of this method of racial differentiation, were not as useful as expected. As may be seen from Table 3 the Baganda Bantomorphs have a very similar blood group gene distribution to the Batutsi Hamitomorphs, while the Bahutus, also a Bantomorphic tribe, have a blood group gene distribution very different from the Baganda.

TABLE 3. DISTRIBUTION OF THE PERCENTAGE SICKLAEMIA AND BLOOD GROUP GENES OF THE VARIOUS EAST AFRICAN RACIAL GROUPS ACCORDING TO LEHMANN AND RAPER (1949), ELSDON-DEW (1950), ALLISON *et al* (1952), AND HIERNAUX (1954)

racial group	tribe	sick-laemia	blood group genes		
			A	B	O
(1)	(2)	(3)	(4)	(5)	(6)
Hamitomorph	Bahima	2.4	11.3	7.1	81.3
	Batutsi	1.4	12.6	5.6	81.7
Bantomorph	*Baganda*	19.0	10.2	6.6	82.9
	Bahutu				
	(Uganda)	8.0	—	—	—
	(Ruanda)	5.2	17.4	13.7	68.7
	(Urundi)	11.8	17.4	13.7	68.7
	Akikuyu	—	17.7	13.4	68.8
Nilotomorph	Luo	28.0	14.9	16.3	68.8
	Lango	27.0	16.3	17.1	64.1
Nilohamitomorph	Masai	—	11.5	17.6	71.0
	Teso	17.8	13.4	13.0	72.2
Congomorph	Baamba	45.0	18.1	12.8	69.1

2.3. Hiernaux (1954) found that the Rhesus-factor incidence in both the Hamitomorphic Batutsi and Bantomorphic Bahutu showed no significant difference.

The same author also discovered that the incidence of sicklaemia in the Ruanda Bahutu was less than half of the metrically very similar Urundi Bahutu.

TABLE 4. RHESUS FACTOR INCIDENCE IN THE BATUTSI AND BAHUTU TRIBES

factor	Batutsi	Bahutu
(1)	(2)	(3)
R.0	63.3	68.0
R.1	6.9	4.7
R.2	5.6	5.9
R′	2.2	1.8
r	21.7	19.3

2.4. This conflicting serological data would appear to emphasise the greater consistency and definition of the somatometrical approach to problems of racial taxonomy. Although Boyd (1947) has stated that somatometry is invalid because it is haphazard in its choice of characters, and that the genetical background of these characters is not understood, these morphometric characteristics have been successfully employed statistically to differentiate between racial groups and their taxonomic value cannot be discounted simply because their inheritance is unknown. In particular, the statistical technique known as Mahalanobis's Generalised Distance, or D^2, (Mahalanobis, 1936), has proved most useful in problems of racial discrimination. In order, therefore, to further the investigation into the relationships between the various East African racial groups one of the authors, D. A. East, has employed this method, as applied by Mahalanobis and Rao to some Indian anthropometric data (Mahalanobis *et al*, 1949), to analyse the results of the survey.

3. D^2 ANALYSIS OF L. OSCHINSKY'S EAST AFRICAN DATA

3.1. The measurements on male adult members of various East African tribes on which the following analysis was carried out have already been the subject of a preliminary report (Oschinsky, 1954). Out of the 18 body and head absolute measurements taken in nearly all cases, the following 14 were considered to be of the greatest importance from the anthropological viewpoint (abbreviations indicated in brackets):

Body: Stature (St), Sitting Height (SH), Sitting Trunk Height (STH), Total Upper Extremity Length (TUL), Total Lower Extremity Length (TLL).

Head: Head Length (HL), Head Breadth (HB), Bizygomatic Breadth (BzB), Interorbital Breadth (IoB), Nose Breadth (NB), Head Height (HH), Morphological Face Height (MFH), Nose Height (NH), Lip Height (LH).

3.2. Unfortunately it was not possible to use all the data compiled in the investigation as, for several tribes, an insufficient number of individuals had been measured to form adequate samples. In fact, there were only eight tribes with sample size, n, greater than 40 for all 14 characters. Although a sample size as small as 40 cannot be considered satisfactory—Mahalanobis and Rao recommend that n should be in the region of 200—a study of the coefficients of variation from the smaller samples showed them to be comparable with those from the larger ones. For the sake of uniformity and ease of computation only individuals for whom all 14 measurements were present were included. Table 5 shows the tribes, comprising four racial groups, and their sample sizes.

TABLE 5. TRIBES COMPRISING FOUR RACIAL GROUPS AND THEIR SAMPLE SIZES

race (1)	tribe (2)	n (3)
Bantomorphs	Baganda	301
	Bahutu	98
	Batoro	57
	Bahiru	44
	Banyoro	43
Hamitomorphs	Batutsi	104
Mulattomorphs	Swahili	104
Nilotomorphs	Luo	64
	total (8 tribes)	815

4. Normality of the Samples

4.1. In order to test for normality of sampling, frequency distributions were prepared, by tribes, for each character, and the statistics g_1 and g_2 worked out together with their standard errors. g_1, which tests a sample for skewness, and g_2 which tests for kurtosis, should both be zero for the distribution of a sample to be considered completely normal. Both g_1 and g_2 can be tested for significance as t with degrees of freedom ∞, i.e., if either g divided by its standard error gives a value in excess of 1.96 then the distribution of the sample concerned is considered to depart from normality at the 5% level of significance ($P < .05$). A positive value of g_1 shows that the distribution is asymmetrical in that the value of the mean is greater than the mode, indicating an excess of measurements less than the mean of the sample; while positive g_2 indicates that the form of the distribution is peaked, i.e., that the sample contains an excess of values very close to the mean: and *vice versa*.

4.2. The 112 sample values of g_1 are shown in Table 6a, from which it may be seen that there was a preponderance of positive values, the ratio of 69 positive to 43 negative values being highly significantly different from the expected 56 each. This overall tendency towards samples containing an excess of low readings was most marked in the Batoro, Swahili and Batutsi tribes, each of which had over twice as many positive as negative values. As regards the various characters, NH had all 8 values positive, TLL had 7, and IoB and MFH each had 6.

4.3. Table 6b shows the values of g_2 for all 112 samples and in this case there was no significant difference between the expected 56 each and the observed ratio of 55 positive and 57 negative values. Two tribes, the Baganda and the Bahutu, each had over twice as many peaked as flat distributions among the 14 characters, while the Swahili had 9 flat and 5 peaked. Of the characters, TLL had 7 peaked and 1 flat, TUL had 6 peaked and 2 flat, while BzB had 6 flat and 2 peaked distributions.

4.4. Thus in general it will be seen from these tables that the two characters in which departure from sample normality was most marked were TLL, which for 6 out of 8 tribes had distributions both positively skew and peaked, and NH, which was always positively skew, with 3 peaked and 5 flat distributions.

4.5. The occurrence of values of g_1 and g_2 which differed from zero at the 5%, or greater, level of significance, is shown in Figure 2.

	ST	SH	STH	TUL	TLL	HL	HB	BZB	IOB	NB	HH	MFH	NH	LH	TOT.
BAGANDA						↑→					↓		←		3
BAHUTU		↑			↑	↑					↑	←		↑	6
BATORO														↑	1
BAHIRU															0
BANYORO														↑→	1
BATUTSI															0
SWAHILI				←									←↑		2
LUO	←	↓		←↓	←						←				5
TOTALS	1	2	0	2	2	2	0	0	0	0	3	1	2	3	18

Fig. 2. Incidence of significantly non-normal samples

← indicates positively skew distribution
→ ,, negatively ,, ,,
↑ ,, peaked ,,
↓ ,, flat ,,

Of the 10 significant values of g_1, against approximately 6 which might have been expected to arise by chance, 8 were positive in sign,—of which 4 occurred in the Luo tribe,—thus supporting the indication that sampling bias was towards observations less than the mean. 12 significant values of g_2 are shown and this is about double the number which might have been expected by chance. Of these 12, 9 were positive, of which 5 came from the Bahutu tribe. Thus, although the 112 samples showed no significant difference between the numbers of peaked and flat distributions, three-quarters of those for which g_2 was significant were peaked.

TABLE 6a. VALUES OF g_1 (MEASURE OF SKEWNESS) AND STANDARD ERRORS

tribe	n	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH	standard errors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Baganda	301	−01.922	−0.0005	+0.0141	−0.0701	−0.1194	−0.3407	−0.2626	−0.0052	+0.2008	−0.0173	−0.1012	+0.0649	+0.5627	+0.0873	0.1405
Bahutu	98	−0.0480	−0.3991	−0.3838	−0.1725	+0.1582	+0.2373	−0.2441	+0.0150	−0.0092	+0.1122	−0.1105	+0.5376	+0.2647	−0.1454	0.2438
Batoro	57	+0.5750	+0.4940	+0.2242	+0.1424	+0.3670	−0.5839	+0.1438	−0.3844	+0.1911	+0.2314	+0.0707	+0.5010	+0.2431	+0.0948	0.3163
Bahiru	44	−0.1020	+0.0043	−0.3334	−0.1390	+0.1414	−0.1480	+0.5894	−0.2770	+0.1395	+0.1372	+0.3586	−0.0696	+0.6867	−0.3675	0.3575
Banyoro	43	+0.0048	−0.2798	+0.0866	+0.0839	+0.2594	+0.2824	+0.5970	−0.1716	−0.2726	+0.1260	−0.1474	+0.4669	+0.1797	−0.7830	0.3614
Batutsi	104	−0.0595	+0.2679	+0.0406	−0.0896	+0.0343	+0.0547	+0.0144	+0.0784	+0.1907	+0.1356	+0.0066	−0.0532	+0.1273	−0.3276	0.2368
Swahili	104	+0.3106	−0.0247	+0.1663	+0.5965	+0.2511	−0.0675	+0.3059	+0.2127	+0.2939	−0.1415	+0.2686	+0.1628	+0.8676	+0.3201	0.2386
Luo	64	+0.6359	−0.2993	−0.1964	+0.6812	+0.6056	−0.0206	−0.2280	+0.1608	+0.5094	−0.0062	+0.7802	+0.3531	+0.0243	+0.0064	0.2993

TABLE 6b. VALUES OF g_2 (MEASURE OF KURTOSIS)

tribe	n	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH	standard errors
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Baganda	301	+0.0164	−0.0654	−0.2747	+0.0118	+0.3037	+1.2873	+0.5416	+0.1794	+0.2122	+0.0376	−0.5925	+0.0356	+0.1836	−0.1114	0.2801
Bahutu	98	+0.4414	+1.6083	+0.7449	+0.0989	+0.9774	+1.0622	−0.4623	−0.3017	+0.1476	−0.4310	+0.9636	+0.5336	−0.2699	+1.4211	0.4830
Batoro	57	+0.5233	−0.6927	+1.0362	+0.1436	+0.7597	−0.3453	+0.0271	−0.0818	−0.3844	+0.7557	−0.2811	−0.2259	−0.4756	+1.2386	0.6086
Bahiru	44	−0.9381	−0.8604	+0.0516	−0.3018	+0.2035	−0.8739	+0.0374	−0.7988	−0.5128	−0.5262	−0.0596	−0.4355	−0.2826	+0.0871	0.7017
Banyoro	43	−0.4400	+0.1840	−0.8837	−0.1225	+0.2877	+0.0424	+0.9959	−0.6827	−0.5737	+0.4564	−0.6121	−0.1019	−0.2268	+1.5884	0.7090
Batutsi	104	+0.0069	+0.0471	−0.1481	+0.5210	+0.4419	−0.7721	−0.3807	+0.3019	−0.1434	−0.4072	+0.1579	−0.1018	−0.0986	+0.9131	0.4695
Swahili	104	−0.1216	−0.3519	−0.6969	+0.0124	−0.6740	+0.0968	−0.1977	−0.6732	+0.7264	−0.4289	+0.7085	−0.2835	+1.3322	−0.2763	0.4695
Luo	64	+1.0082	−1.3946	−0.3436	+1.3446	+0.8228	−0.2340	−0.0620	−0.2726	−0.5351	−0.6359	+1.0441	+0.8040	+0.3433	−0.3081	0.5905

4.6. No case of significant departure from normality of sampling was found for either the Bahiru or Batutsi tribes and only one, LH, for both Batoro and Banyoro. Five characters, STH, HB, BzB, IoB and NB, gave rise to no abnormal distributions, but characters HH and LH each had three distributions significantly different from normal. The occurrence of significant departure from normality was apparently unaffected by the size of the sample.

4.7. The probabilities of obtaining values less than, or equal to, the observed values of g_1 and g_2 were calculated and distributed by decile range. The results are shown in Table 7 columns (2) and (3).

TABLE 7. FREQUENCY DISTRIBUTIONS OF THE PROBABILITY OF OBTAINING VALUES OF g_1 AND g_2, EQUAL TO OR LESS THAN THOSE OBSERVED

decile range (1)	g_1 (2)	g_2 (3)
0.0–0.1	8	7
0.1–0.2	6	9
0.2–0.3	9	12
0.3–0.4	10	14
0.4–0.5	10	10
0.5–0.6	14	12
0.6–0.7	11	11
0.7–0.8	14	6
0.8–0.9	10	6
0.9–1.0	20	25
	112	112

Theoretically these frequency distributions should be rectangular, i.e., if normally distributed one-tenth of the 112 probabilities should lie within each decile range. For g_1 a χ^2 test of the observed against the expected frequencies was insignificant, $\chi^2 = 12.688$ with 9 degrees of freedom, and so the probabilities of occurrence of the g_1 values could be said to be normally distributed in spite of the high figure of 20 for the range 0.9—1.0. In the case of g_2, $\chi^2 = 27.910$, significant at the 0.1% point ($P < .001$), showing that the probabilities of occurrence of the g_2's were not normally distributed, the greatest contribution to χ^2 again coming from the range 0.9—1.0, with a frequency of 25 against an expected 11.2, indicating that the distribution of g_2 values was abnormal in that there was a preponderance of extreme values.

5. Sample means and standard errors

For each tribe, Table 9 shows the means and their standard errors for each of the 14 characters. Owing to the amount of computation involved, the only t-tests carried out were those on the differences, for each character, between the means of the Baganda and those of the other seven tribes[1]. The significant differences arising from these tests are shown in Table 8.

TABLE 8. RESULTS OF t-TESTS ON THE DIFFERENCES BETWEEN BAGANDA MEANS AND THOSE OF THE OTHER SEVEN TRIBES. S INDICATES SIGNIFICANT DIFFERENCE AT, OR ABOVE, THE 5% LEVEL ($P < .05$)

	body characters					head characters								
Baganda with	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Bahutu	S	—	—	—	—	S	—	S	—	S	S	S	—	—
Batoro	S	S	S	—	—	S	—	—	—	S	S	—	S	—
Bahiru	—	—	—	—	—	S	S	S	—	S	S	—	—	—
Banyoro	—	—	—	—	—	S	—	S	—	—	S	S	—	—
Batutsi	S	S	S	S	S	S	—	S	S	S	—	—	S	—
Swahili	—	—	—	—	—	S	S	—	—	S	—	S	S	—
Luo	S	S	S	S	S	—	S	S	S	—	—	—	S	—

5.2. The coefficients of variation for each character, by tribes, are shown in Table 10 from which it will be seen that in general there was good agreement within each character. Taking rough averages, St, SH, HL, HB and BzB were less than 4, STH, TUL, TLL, HH and MFH were less than 5, NH was less than 6, NB was less than 7, IoB was just about 8, while LH was about 13.

5.3. For each character the eight sample (tribe) variances were tested for homogeneity by the Bartlett test. Only three characters showed significant heterogeneity, two of these, NH and LH, at the 5% level, and one, IoB, at the 0·1% level.

5.4. *Correlation coefficients.* Correlation coefficients were worked out between all 14 characters for each of the eight tribes. It was noticed that, in general, the body characters were more highly inter-correlated than those of the head; in particular, the highest, in nearly all cases, being those in which St was present.

---

[1] The Baganda tribe was made the basis for these comparisons because (a) it had the largest n, (b) its mean values for the various characters appeared to be fairly centrally situated, and (c) it had been the focal point of interest in the previous examination of the data.

TABLE 9. SAMPLE MEANS AND THEIR STANDARD ERRORS (mm.)

tribe	n	St	SH	STH	TUL	TLL	HL	HB
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Baganda	301	1658.27±3.43	844.46±1.91	548.19±1.61	779.18±2.08	909.32±2.59	195.07±0.37	142.61±0.28
Bahutu	98	1641.45 6.20	842.53 3.35	554.19 2.60	777.45 3.83	901.60 4.80	190.66 0.60	142.82 0.42
Batoro	57	1677.39 8.66	855.67 4.11	560.82 3.43	786.84 4.61	920.96 5.95	192.42 0.77	143.56 0.54
Bahiru	44	1647.36 9.50	838.30 4.82	547.45 3.92	774.20 5.73	904.25 7.02	190.73 0.98	139.32 0.67
Banyoro	43	1664.05 9.72	841.05 4.64	546.09 4.24	771.06 5.71	912.35 7.52	191.77 0.92	143.05 0.75
Batutsi	104	1723.03 5.79	865.67 3.19	563.45 2.72	798.70 3.39	959.79 4.76	193.21 0.56	142.88 0.48
Swahili	104	1649.10 5.80	840.88 2.96	545.62 2.32	771.99 3.68	900.06 4.19	191.80 0.69	147.22 0.55
Luo	64	1733.30 6.83	870.30 3.84	562.14 3.42	803.27 4.15	959.11 5.07	194.56 0.86	147.22 0.64

TABLE 9.—(*continued*)

tribe	n	BzB	IoB	NB	HH	MFH	NH	LH
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Baganda	301	136.21±0.32	36.23±0.17	44.90±0.16	122.37±0.34	118.34±0.32	49.49±0.15	25.88±0.20
Bahutu	98	134.71 0.48	35.82 0.28	43.35 0.30	117.54 0.55	116.30 0.68	49.01 0.28	26.27 0.34
Batoro	57	137.02 0.61	36.26 0.43	43.86 0.31	119.86 0.75	118.79 0.70	51.19 0.31	26.51 0.44
Bahiru	44	132.59 0.69	36.11 0.38	43.52 0.43	118.70 0.91	116.64 0.85	49.16 0.46	25.30 0.41
Banyoro	43	133.86 0.82	36.02 0.48	44.40 0.43	119.49 0.89	115.30 0.87	49.14 0.36	25.44 0.50
Batutsi	104	132.58 0.53	34.41 0.29	39.58 0.27	122.88 0.54	118.36 0.56	51.95 0.32	26.29 0.27
Swahili	104	135.43 0.52	35.75 0.21	44.08 0.29	121.40 0.54	116.20 0.56	51.69 0.25	26.01 0.31
Luo	64	138.23 0.58	37.58 0.42	45.56 0.29	123.07 0.62	118.17 0.76	51.11 0.31	26.83 0.51

TABLE 10. COEFFICIENTS OF VARIATION

(standard error of the sample expressed as % of the sample mean)

tribe	n	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Baganda	301	3.59	3.92	5.10	4.63	4.94	3.28	3.37	3.94	8.07	6.41	4.79	4.75	5.13	13.72
Bahutu	98	3.74	3.93	4.64	4.88	5.27	3.14	2.87	3.53	7.76	6.94	4.66	5.81	5.74	12.61
Batoro	57	3.90	3.62	4.62	4.42	4.87	3.02	2.85	3.38	8.91	5.41	4.70	4.43	4.54	12.50
Bahiru	44	3.87	3.82	4.75	4.91	5.15	3.40	3.19	3.43	7.07	6.49	5.67	4.81	6.23	16.77
Banyoro	43	3.83	3.61	4.24	4.85	5.41	3.15	3.42	4.01	8.79	6.34	4.86	4.94	4.86	12.96
Batutsi	104	3.43	3.76	4.92	4.33	5.06	2.94	3.39	4.07	8.67	7.01	4.50	4.81	6.25	10.64
Swahili	104	3.59	3.59	4.33	4.86	4.75	3.67	3.83	3.88	5.90	6.81	4.53	4.91	5.01	12.25
Luo	64	3.15	3.53	4.87	4.14	4.22	3.55	3.47	3.33	9.06	5.14	4.05	5.12	4.87	15.12

6. Elimination of certain characters

6.1. At this point it was considered desirable to attempt the selection of a reduced list of characters on which to base the final analysis. It was decided that about 9 of the original selection of 14 characters should be retained, as more than this number would greatly increase the amount of computation involved in the later stages.

6.2. Although Table 8, which shows the results of the *t*-tests carried out, considered in conjunction with the results of the tests on normality of sampling, homogeneity of variances, correlation coefficients, etc., might have been used to form the basis for such a selection, a further attempt, based on the evaluation of successive D^2 values, was made to determine which of the 14 characters were most valuable in differentiating between the tribes.

6.3. As in the *t*-tests[1] the Baganda figures were again taken, with those of the other tribes and Table 11 shows, for body and head characters separately, the significance level reached by D^2 as each character was successively introduced.

TABLE 11. SIGNIFICANCE LEVEL REACHED BY D^2 IN COMPARING THE BAGANDA WITH EACH OF THE OTHER SEVEN TRIBES, FOR BODY AND HEAD SEPARATELY, AS EACH CHARACTER SUCCESSIVELY INTRODUCED

	body characters					head characters								
Baganda with	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Bahutu	*	*	***	***	***	***	***	***	***	***	***	***	***	***
Batoro	*	*	**	**	*	**	***	**	**	**	***	***	***	***
Bahiru			*			***	***	***	***	***	***	***	***	***
Banyoro			*	**	**	**	**	***	***	***	***	***	***	***
Batutsi	***	***	***	***	***	**	*	***	***	***	***	***	***	***
Swahili						***	***	***	***	***	***	***	***	***
Luo	***	***	***	***	***		***	***	***	***	***	***	***	***

* indicates D^2 significant at 5% level ($P < .05$)
** " D^2 " " 1% " ($P < .01$)
*** " D^2 " " 0.1% " ($P < .001$)

---

[1] As any one *t*-test may be said to indicate the significance of the distance between the Baganda and another tribe in respect of a single character it is equivalent to D^2 in its simplest form. This is shown by comparison of the *t* and D^2 results for St and HL, *i.e.*, the first characters used for body and head respectively.

It was then possible to prepare a further table (Table 12) showing, again for body and head characters separately, the significance of the increase in ability to differentiate due to bringing in each additional character.

TABLE 12. SIGNIFICANCE OF THE INCREASE IN D^2 IN COMPARING THE BAGANDA WITH EACH OF THE OTHER SEVEN TRIBES, FOR BODY AND HEAD SEPARATELY, DUE TO THE INTRODUCTION OF EACH SUCCESSIVE CHARACTER

	body characters					head characters								
Baganda with	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Bahutu	*		***	**		***	*	**		**	***			***
Batoro	*		**			**	*			*	*		***	
Bahiru			*			***	**				*			
Banyoro			**	**		**		**			**	*		
Batutsi	***			***		**		***	**	***			***	
Swahili						***	***	***			**	*	***	
Luo	***		*	***			***		**				***	*

* indicates increase in D^2 significant at 5% level
** ,, ,, ,, D^2 ,, ,, 1% ,,
*** ,, ,, ,, D^2 ,, ,, 0.1% ,,

From the above table it will be seen that two body characters, SH and TLL, and three head, IoB, MFH and LH, were comparatively ineffective in aiding discrimination between the Baganda and the majority of the other tribes. This provides corroboration of any selection of characters based on Table 8, etc. As, however, the importance of LH had been stressed from the anthropological viewpoint this was retained in some further calculations.

7. Pooled estimates of correlation

7.1. Having thus reduced the effective body characters to three, St, STH and TUL, and the head characters to seven, HL, HB, BzB, NB, HH, NH and LH, the 45 pooled estimates of correlation coefficients between these 10 characters were obtained, by use of the r to z transformation, and tested for significance. These are shown in Tables 13a and 13b together with the results of χ^2 tests for the homogeneity of each set of eight coefficients on which the pooled estimates were based.

7.2. From Table 13a and Table 13b it may be seen firstly, that all the pooled correlation coefficients were positive in sign, and secondly, that of the total of 45 over half, 26, were less than 0.2, and 42 were less than 0.4, in magnitude; two

TABLE 13a. POOLED CORRELATION COEFFICIENTS AND THEIR SIGNIFICANCE

	St	STH	TUL	HL	HB	BzB	NB	HH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
St		+.5949***	+.7888***	+.3479***	+.2240***	+.3040***	+.1810***	+.2316***	+.2084***	+.1303***
STH			+.3785***	+.2590***	+.1625***	+.2002***	+.1116**	+.0997**	+.1704***	+.0968**
TUL				+.2552***	+.1548***	+.2887***	+.1791***	+.1450***	+.1733***	+.1075**
HL					+.3408***	+.3104***	+.2416***	+.3004***	+.0779*	+.1312***
HB						+.5475***	+.1293***	+.2664***	+.0340	+.0438
BzB							+.3022***	+.1743***	+.0460	+.0591
NB								+.1037**	+.1040**	+.0405
HH									+.0283	+.1988***
NH										+.0498

* indicates significance at 5% level
** ,, ,, ,, 1% ,,
*** ,, ,, ,, .1% ,,

TABLE 13b. χ^2 VALUES RESULTING FROM TESTS FOR HOMOGENEITY OF CORRELATION COEFFICIENTS FROM WHICH THE POOLED COEFFICIENTS (a) WERE DERIVED

	St	STH	TUL	HL	HB	BzB	NB	HH	NH	LH
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
STH	4 03									
TUL	7.69	10.14								
HL	14.17*	6.20	20.54**							
HB	16.17*	10.55	4.00	13.78						
BzB	15.23*	11.24	5.50	13.02	12.12					
NB	12.84	6.05	6.47	4.30	6.99	9.85				
HH	5.64	10.14	6.41	10.70	3.69	2,64	10.38			
NH	12.22	6.89	11.69	7.70	5.48	6.40	9.02	9.63		
LH	16.22*	16.77*	4.22	4.70	6.91	4.10	9.23	15.06*	8.11	

* indicates significance at 5% level ($\chi^2 = 14.07$)
** ,, ,, ,, 1% ,, ($\chi^2 = 18.48$)
*** ,, ,, ,, .1% ,, ($\chi^2 = 24.32$)

of the three remaining values were within the range 0.5-0.6 and only one, St with STH, had the comparatively high value of +0.7888. As stated earlier, the highest values were those between pairs of body characters, in particular those in which St was present, the most outstanding exception to this being the value +0.5475 for the head characters HB with BzB. Conversely, in general, the lowest pooled correlations were those involving LH, though the smallest of all was that for HH with NH with a value of only +0.0283.

7.3. That homogeneity existed within each set of eight individual correlations was disproved in only 7 cases out of 45, of which 6 were at the 5% level and 1 only, TUL with HL, at the 1% level of significance. In those 6 sets of coefficients in which heterogeneity was found to be significant at the 5% level, St was involved 4 times, with HL, HB, BzB and LH; and LH was involved 3 times, with St, STH and HH.

8. NORMALISED MEAN VALUES

By subtracting the weighted character mean from each tribal mean and dividing by the pooled estimate of the standard deviation for that character, normalised mean values were obtained and these are shown in Table 14.

TABLE 14. NORMALISED MEAN VALUES

tribe (1)	St (2)	STH (3)	TUL (4)	HL (5)	HB (6)	BzB (7)	NB (8)	HH (9)	NH (10)	LH (11)
Baganda	−0.265	−0.200	−0.101	+0.401	−0.201	+0.220	+0.440	+0.302	−0.317	−0.056
Bahutu	−0.545	+0.026	−0.149	−0.295	−0.158	−0.072	−0.110	−0.552	−0.496	+0.059
Batoro	+0.052	+0.276	+0.111	−0.017	−0.004	+0.377	+0.071	−0.141	+0.317	+0.131
Bahiru	−0.447	−0.228	−0.239	−0.284	−0.883	−0.484	−0.050	−0.347	−0.440	−0.229
Banyoro	−0.169	−0.279	−0.327	−0.120	−0.110	−0.237	+0.263	−0.207	−0.447	−0.187
Batutsi	+0.811	+0.376	+0.439	+0.107	−0.145	−0.486	−1.449	+0.393	+0.600	+0.065
Swahili	−0.418	−0.297	−0.300	−0.115	+0.755	+0.068	+0.149	+0.131	+0.503	−0.018
Luo	+0.981	+0.326	+0.565	+0.320	+0.755	+0.613	+0.675	+0.426	+0.287	+0.226

These figures indicate, for each character, the mean difference per unit standard deviation between each tribe and the overall mean and thus, in effect, all the figures in the table have been reduced to the same 'scale' and are therefore comparable in size, the original inequalities in magnitude between the various character measurements having been eliminated.

9. D^2 VALUES

9.1. Table 15 below shows the mean values of the transformed characters obtained by the method explained in detail by Rao (1952). These are such that the generalised distance, or D^2, between any two tribes is the sum of squares of the differences between their corresponding y-values.

TABLE 15. MEAN VALUES OF TRANSFORMED CHARACTERS

tribe	y_1	y_2	y_3	y_4	y_5	y_6	y_7	y_8	y_9	y_{10}
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Baganda	−0.265	−0.052	+0.169	+0.534	−0.303	+0.382	+0.323	+0.319	−0.289	−0.123
Bahutu	−0.545	+0.435	+0.546	−0.130	0	+0.082	−0.043	−0.360	−0.435	+0.199
Batoro	+0.052	+0.305	+0.172	−0.056	−0.008	+0.439	−0.041	−0.113	+0.301	+0.134
Bahiru	−0.447	+0.047	+0.197	−0.136	−0.794	+0.027	+0.076	−0.076	−0.384	−0.154
Banyoro	−0.169	−0.221	−0.362	−0.058	−0.063	−0.135	+0.386	−0.204	−0.467	−0.109
Batutsi	+0.811	−0.132	−0.358	−0.185	−0.302	−0.648	−1.475	+0.307	+0.559	−0.086
Swahili	−0.418	−0.060	+0.038	+0.037	+0.903	−0.274	+0.259	+0.072	+0.599	+0.021
Luo	+0.981	−0.321	−0.406	−0.009	+0.576	+0.125	+0.508	+0.062	+0.089	+0.110

The D^2 values thus obtained, excluding LH (y_{10}), are shown in Table 16a arranged in ascending order.

9.2. When tested as $\chi^2 = n_1 n_2 D^2/(n_1+n_2)$, with degrees of freedom equal to the number of characters used, i.e., 9, all were significant at the 0.1% point with the sole exception of Bahiru with Banyoro which was significant at the 1% point. This indicates that the nine characters chosen were sufficient to ensure a very high degree of differentiation between all the tribes, including those within the Bantomorphic group. In fact the level of significance of the D^2 values was, in general, so high that, had a somewhat greater number of observations been taken on those Bantomorphic tribes with very low n, it is probable that fewer than nine characters might justifiably have been used with successful result.

It will clearly be seen from Table 16a and Figure 3 that the Bantomorphic tribes were closely clustered together, and that this cluster was progressively distant from the Mulattomorphs (Swahili), Nilotomorphs (Luo), and the Hamitomorphs

TABLE 16a. D^2 VALUES IN ASCENDING ORDER EXCLUDING LH

Baganda	*with*	Bahutu	*with*	Batoro	*with*	Bahiru	*with*	Banyoro	*with*	Batutsi	*with*	Swahili	*with*	Luo	*with*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Bahiru	1.086	Bahiru	1.013	Bahutu	1.249	Bahutu	1.013	Bahiru	1.144	Batoro	4.632	Batoro	1.932	Banyoro	2.174
Banyoro	1.294	Batoro	1.249	Baganda	1.334	Baganda	1.086	Baganda	1.294	Luo	5.680	Banyoro	2.373	Batoro	2.408
Batoro	1.334	Banyoro	1.662	Bahiru	1.594	Banyoro	1.144	Bahutu	1.662	Banyoro	6.023	Luo	2.813	Swahili	2.813
Bahutu	1 697	Baganda	1.697	Banyoro	1.678	Batoro	1.594	Batoro	1.678	Bahiru	6.064	Bahutu	2.836	Baganda	3.332
Swahili	3 026	Swahili	2.836	Swahili	1.932	Swahili	4.060	Luo	2.174	Swahili	6.377	Baganda	3.026	Bahiru	4.870
Luo	3.332	Luo	4.911	Luo	2.408	Luo	4.870	Swahili	2.373	Baganda	6.972	Bahiru	4.060	Bahutu	4.911
Batutsi	6.972	Batutsi	7.088	Batutsi	4.632	Batutsi	6.064	Batutsi	6.023	Bahutu	7.088	Batutsi	6.377	Batutsi	5 680

TABLE 16b. D^2 VALUES IN ASCENDING ORDER INCLUDING LH

Baganda	*with*	Bahutu	*with*	Batoro	*with*	Bahiru	*with*	Banyoro	*with*	Batutsi	*with*	Swahili	*with*	Luo	*with*
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Bahiru	1.087	Bahiru	1.138	Bahutu	1.253	Baganda	1.087	Bahiru	1.146	Batoro	4.680	Batoro	1.945	Banyoro	2.222
Banyoro	1.294	Batoro	1.253	Baganda	1.400	Bahutu	1.138	Baganda	1.294	Luo	5.718	Banyoro	2.390	Batoro	2.409
Batoro	1.400	Banyoro	1 757	Bahiru	1.677	Banyoro	1.146	Batoro	1.737	Banyoro	6.024	Luo	2.821	Swahili	2.821
Bahutu	1.801	Baganda	1.801	Banyoro	1.737	Batoro	1.677	Bahutu	1.757	Bahiru	6.069	Bahutu	2.868	Baganda	3.386
Swahili	3.047	Swahili	2.868	Swahili	1.945	Swahili	4.091	Luo	2.222	Swahili	6.388	Baganda	3.047	Bahutu	4.919
Luo	3.386	Luo	4.919	Luo	2.409	Luo	4.940	Swahili	2.390	Baganda	6.973	Bahiru	4.091	Bahiru	4.940
Batutsi	6.973	Batutsi	7.169	Batutsi	4.680	Batutsi	6.069	Batutsi	6.024	Bahutu	7.169	Batutsi	6.388	Batutsi	5.718

(Batutsi). Of the three last-mentioned groups the nearest together were the Mulattomorphs and Nilotomorphs. The intra- and inter-cluster average D^2 values are shown in Table 17.

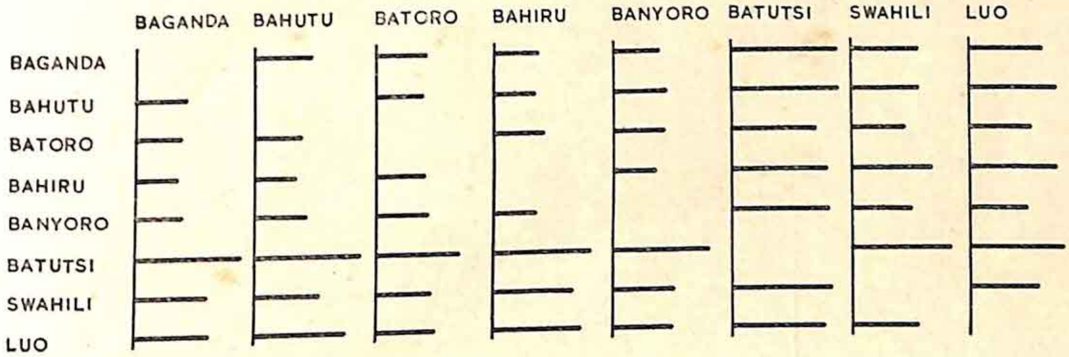


Fig. 3. Generalised distances (D's) between tribes.

TABLE 17. INTRA AND INTER-CLUSTER AVERAGE D^2 VALUES

racial group	Banto-morphs	Mulatto-morphs	Niloto-morphs
Bantomorphs	1.375		
Mulattomorphs	2.864		
Nilotomorphs	3.482	2.813	
Hamitomorphs	6.111	6.377	5.680

Taking the square-roots of these D^2's to represent the distances separating the clusters, Figure 4 shows their approximate mutual relationships.

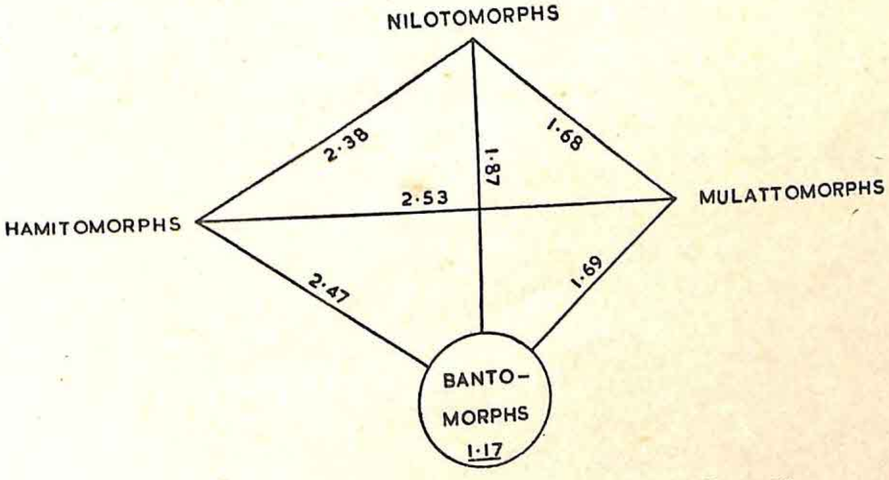


Fig. 4. Approximate mutual relationships of the racial groups.

It is interesting to see how closely these approximate relationships between the racial groups agree with the general geographical locations of the groups as given by Oschinsky (1954) and reproduced in Figure 5.

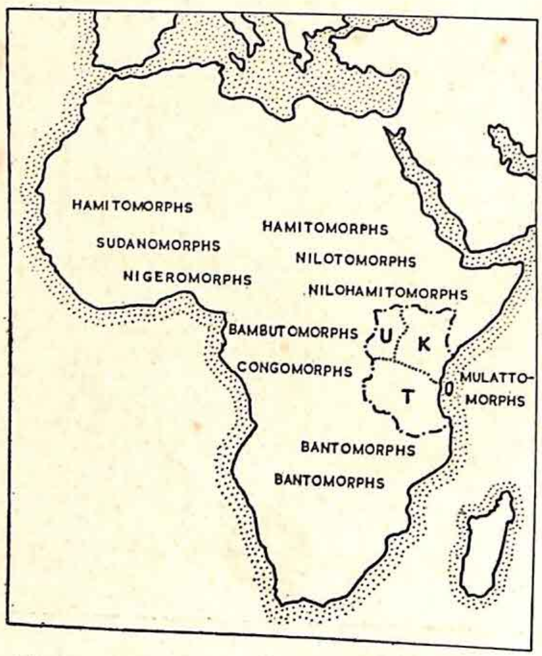


Fig. 5. Distribution of the negro and negroid races of Africa.

Key :

U = Uganda
K = Kenya
T = Tanganyika

10. EFFECT OF BRINGING IN AN ADDITIONAL CHARACTER

Although it was proved that the nine characters selected primarily for statistical reasons allowed for very effective discrimination between the tribes, it was decided to see what effect would be obtained by the addition of head character LH. The normalised mean values and the mean values of the transformed characters are shown in Tables 14 and 15 and the new values of D^2 in Table 16b. It will be seen from a comparison of Tables 16a and 16b, that the effect of introducing LH was very slight. In fact the effect was so small in the case of Bahiru with Banyoro that the χ^2 value, based on 10 instead of 9 characters, was even further from reaching the 0.1% level of significance than before. One result of bringing LH into the analysis was, however, slightly to alter the arrangement of the Bantomorphic tribes in order of increasing distance between them. Table 18 shows this order before and after addition of LH.

TABLE 18. BANTOMORPHIC TRIBES ARRANGED IN ORDER OF INCREASING D^2

excluding LH (1)	D^2 (2)	including LH (3)	D^2 (4)
Bahutu—Bahiru	1.013	Baganda—Bahiru	1.087
Baganda—Bahiru	1.086	Bahutu—Bahiru	1.138
Banyoro—Bahiru	1.144	Banyoro—Bahiru	1.146
Bahutu—Batoro	1.249	Bahutu—Batoro	1.253
Baganda—Banyoro	1.294	Baganda—Banyoro	1.294
Baganda—Batoro	1.334	Baganda—Batoro	1.400
Batoro—Bahiru	1.594	Batoro—Bahiru	1.677
Bahutu—Banyoro	1.662	Banyoro—Batoro	1.737
Banyoro—Batoro	1.678	Bahutu—Banyoro	1.757
Baganda—Bahutu	1.697	Baganda—Bahutu	1.801

The average intra-cluster distance between tribes of the Bantomorphic group was increased from 1.17 to 1.20, and the inter-cluster average distances for the Bantomorphic with Nilotomorphic and Hamitomorphic groups were both increased by 0.01, to 1.89 and 2.49 respectively. The other relationships remained unchanged.

11. CONTRIBUTIONS OF THE VARIOUS CHARACTERS

In order to study further the contributions of the various characters used, their normalised mean values were plotted in Figure 6.

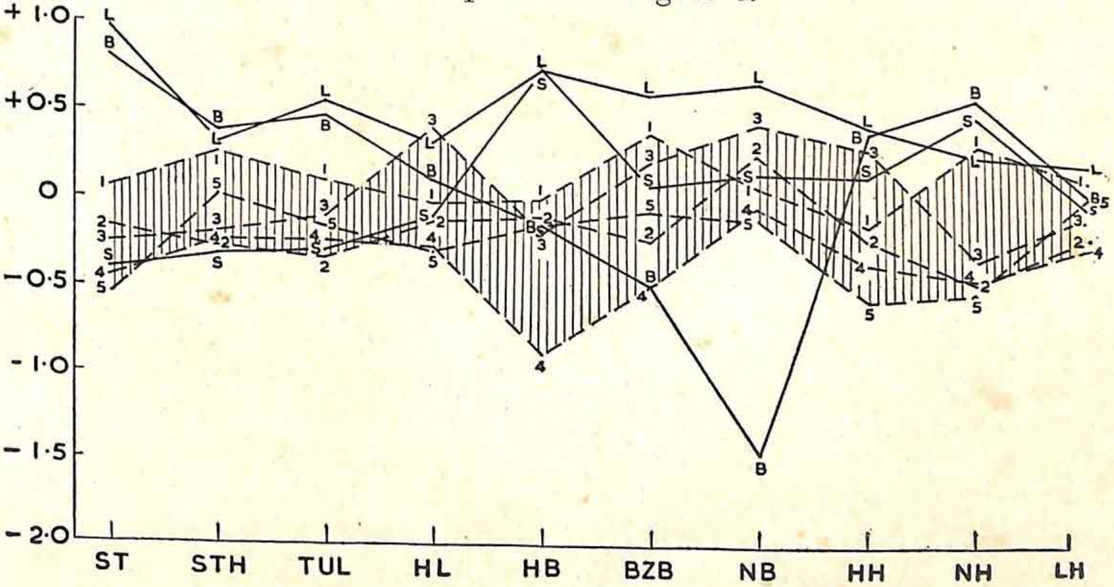


Fig. 6. Normalised mean values of characters used in D^2.

Key :

L = Luo
B = Batutsi
S = Swahili

1 = Batoro
2 = Banyoro
3 = Baganda
4 = Bahiru
5 = Bahutu

The Figure shows that in general, among the Bantomorphic tribes, the Batoro and Baganda had values above, and Bahutu and Bahiru had values below, the average; the greatest degree of divergence in this group being in Head and Nose Height, and Head and Bizygomatic Breadth. As regards the other three racial groups the Nilotomorphs (Luo) had the highest values for most characters; the Mulattomorphs(Swahili) had comparatively low values for body, but high values for head characters, in particular Head Breadth; while the Hamitomorphs (Batutsi) had generally high values with the exceptions of Nose Breadth, for which character they were outstandingly removed from all the other tribes, and Head and Bizygomatic Breadth. The characters which showed the greatest differences between other groups and the Bantomorphs were:

(a) Nilotomorphs and Hamitomorphs: Stature, Total Upper Extremity Length, Nose Breadth;

(b) Nilotomorphs and Mulattomorphs: Head Breadth;

(c) Hamitomorphs and Mulattomorphs: Nose Height.

12. SUMMARY FINDINGS

A highly significant degree of differentiation between the eight tribes considered was found for values of D^2 based on three body measurements, Stature, Sitting Trunk Height and Total Upper Extremity Length, and six head measurements, Head Length, Head Breadth, Bizygomatic Breadth, Nose Breadth, Head Height and Nose Height. The addition of measurements on Lip Height did not materially affect the result. The Bantomorphic group showed a close affinity and were clearly differentiated from the other three racial groups, being progressively distant from the Mulattomorphs, Nilotomorphs and Hamitomorphs. The Hamitomorphs were further removed from the other three groups than any of these were from each other.

13. ACKNOWLEDGEMENT

The authors wish to acknowledge their indebtedness to Dr. John Wishart, Director of the Statistical Laboratory, for advice and facilities in connection with the statistical work. Miss Patricia Johnson rendered valuable assistance with the extensive computation involved in the analysis.

REFERENCES

ALLISON, A., ILKIN, E., MOURANT, A. AND RAPER, A. (1952): Blood groups in some East African tribes. *J. Roy. Anthropol. Inst.*, 82, 55.

BOYD, W. C. (1947): The use of genetically determined characters, especially serological factors such as *Rh*, in physical anthropology. *South West J. Anthropol.*, 3, 32.

ELSDON-DEW, R. (1950): The ethnology of sicklaemia in Uganda. *Nature*, **165**, 763.

HIERNAUX, J. (1954): Les Caracteres Physiques des Populations du Ruanda et de l'Urundi. *Mem. Inst. Roy. des Sci. Nat. Belg.*, II, No. 52.

JOHNSTON, H. H. (1902): *The Uganda Protectorate.* Hutchinson & Co., London.

LEHMANN, F. AND RAPER, A. (1949): Distribution of the sicklecell trait in Uganda and its ethnological significance. *Nature,* **164**, 494; also *Uganda J.*, **15**, 41

LEYS, N. AND JOYCE, T. (1913): Note on a series of physical measurements from East Africa. *J. Roy. Anthropol. Inst.,* **43**, 195.

MAHALANOBIS, P. C. (1936): On the generalised distance in statistics. *Proc. Nat. Inst. Sci.,* (India), **12**, 49.

MAHALANOBIS, P. C., MAJUMDAR, D. N. AND RAO, C. R. (1949): Anthropometric survey of the United Provinces, 1941: A statistical study. *Sankhyā,* **9**, 90.

OSCHINSKY, L. (1954): *The Racial Affinities of the Baganda and other Bantu Tribes of British East Africa.* W. Heffer and Sons, Cambridge.

RAO, C. R. (1952): *Advanced Statistical Methods in Biometric Research,* John Wiley & Sons, New York.

SELIGMAN, C. G. (1913): Some aspects of the Hamitic problem in the Anglo-Egyptian Sudan. *J. Roy. Anthropol. Inst.,* **43**, 593.

——— (1939) : *Races of Africa,* Thornton, Butterworth, Ltd., London.

Paper received : September, 1955.

APPENDIX: MEASUREMENTS ON EAST AFRICAN MALE ADULTS (mm)

RACIAL GROUP: BANTOMORPH

BAGANDA

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
1	1632	858	571	771	879	196	137	126	36	48	131	119	52	20
2	1674	823	524	776	945	193	142	133	33	43	132	119	53	26
3	1618	821	532	776	887	186	143	139	38	46	119	121	52	27
4	1647	804	513	760	907	191	143	136	35	42	125	119	50	20
5	1660	896	595	773	897	202	147	142	32	44	120	121	49	31
6	1613	781	490	764	886	202	144	138	37	44	120	110	51	28
7	1667	834	532	796	931	194	137	128	36	50	118	123	47	27
8	1612	848	531	753	834	198	149	139	32	49	134	118	49	29
9	1700	859	584	766	930	194	145	135	42	44	124	108	49	23
10	1655	836	539	793	902	193	140	142	41	42	130	129	46	29
11	1730	894	577	826	922	196	152	144	32	44	132	120	47	31
12	1698	862	554	806	945	189	145	136	37	45	129	121	47	27
13	1672	826	534	767	922	200	149	144	36	47	127	116	46	30
14	1646	816	516	774	927	193	140	132	31	44	120	109	49	23
15	1741	914	598	789	917	200	147	140	33	48	125	112	50	26
16	1581	807	517	738	855	197	146	139	39	41	119	117	47	22
17	1613	821	526	734	882	189	144	132	37	47	121	124	53	22
18	1624	846	555	758	885	197	144	132	34	44	126	109	48	21
19	1644	844	537	791	924	189	138	128	32	41	126	120	49	27
20	1680	805	497	805	919	196	147	138	32	50	124	125	50	25
21	1667	830	537	797	927	192	141	139	34	47	117	118	47	30
22	1566	801	510	701	833	185	133	128	36	42	123	109	49	31
23	1728	877	540	790	925	197	140	139	38	47	124	118	49	28
24	1677	816	511	814	969	195	136	131	35	51	124	120	53	25
25	1636	821	537	777	896	204	140	134	35	46	131	115	49	27
26	1762	888	567	837	968	204	143	132	36	40	131	126	47	26
27	1710	834	539	794	960	186	142	136	37	44	132	112	48	26
28	1600	819	510	748	878	185	147	135	31	39	129	113	48	25
29	1638	828	527	746	887	186	147	135	39	44	123	113	49	22
30	1703	834	534	814	977	176	141	134	34	42	122	111	46	24
31	1596	769	483	739	869	196	138	131	35	42	129	120	50	26
32	1689	852	545	800	921	196	148	138	35	48	115	128	49	25
33	1640	848	566	765	897	188	140	134	37	48	125	123	55	29
34	1755	837	530	792	961	202	143	135	36	44	115	116	48	21
35	1585	848	551	731	856	199	146	144	37	45	132	120	50	24
36	1665	807	489	774	925	192	143	132	40	47	133	121	51	32
27	1740	824	520	829	983	208	146	138	36	43	132	123	49	24
38	1612	803	520	793	910	183	143	136	35	37	118	118	48	23
39	1692	881	600	766	928	196	157	142	34	44	120	115	48	27
40	1664	853	549	767	911	196	146	140	32	48	127	119	47	25
41	1709	866	577	816	953	204	143	139	40	44	116	123	51	25
42	1635	851	565	766	882	197	142	139	40	39	121	124	52	26
43	1752	923	623	856	956	201	137	130	37	47	126	112	52	25
44	1603	836	539	731	869	194	134	125	33	41	119	116	48	26
45	1606	852	566	736	849	183	147	140	36	42	111	116	47	22

RACIAL GROUP : BANTOMORPH (*continued*)

BAGANDA (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
46	1677	834	563	754	945	194	146	136	33	44	111	119	49	24
47	1676	843	549	785	915	191	141	138	33	40	113	112	47	27
48	1670	836	542	792	934	199	153	145	36	43	121	116	50	26
49	1739	930	612	809	927	206	151	150	41	47	123	115	51	20
50	1707	802	518	812	953	194	145	145	28	47	123	113	47	23
51	1667	864	573	770	932	197	151	140	42	42	127	120	48	27
52	1708	855	571	796	958	198	145	140	41	42	121	128	53	26
53	1520	767	509	733	843	165	122	122	33	45	109	103	50	22
54	1715	845	556	795	965	199	145	147	42	53	126	121	49	28
55	1729	858	549	848	985	203	147	127	37	42	133	125	48	27
56	1740	837	547	798	1002	197	147	137	33	50	126	121	51	23
57	1724	884	584	830	934	201	144	135	30	45	132	128	55	26
58	1705	862	555	804	934	203	142	142	39	49	123	124	50	32
59	1527	745	479	715	849	192	142	132	39	42	123	116	52	28
60	1679	847	552	780	958	199	144	145	36	48	110	125	47	28
61	1654	781	501	762	922	193	132	130	36	46	117	119	46	26
62	1554	798	524	712	846	182	132	140	42	49	117	115	50	27
63	1768	893	566	856	954	201	147	145	38	471	22	121	47	27
64	1711	848	557	794	936	188	148	138	37	37	127	118	56	35
65	1738	842	555	793	1004	194	141	137	36	49	126	108	49	20
66	1623	802	511	774	892	197	138	131	34	42	121	112	48	30
67	1627	831	547	769	888	198	143	135	36	42	126	113	49	26
68	1695	814	519	760	937	188	143	132	37	41	130	121	49	29
69	1711	822	510	802	988	194	142	127	32	47	125	116	47	24
70	1646	828	550	740	908	194	143	138	36	45	125	120	51	26
71	1772	880	566	844	990	206	146	137	35	44	126	121	50	22
72	1617	832	542	747	881	197	146	131	33	46	121	110	48	20
73	1590	777	498	722	880	188	135	125	35	44	111	115	49	25
74	1746	856	541	773	922	199	147	138	37	48	129	111	51	26
75	1656	836	546	787	929	188	147	146	39	49	127	114	50	21
76	1646	827	515	774	913	189	138	127	31	43	128	125	52	29
77	1629	808	514	805	919	192	145	137	39	45	126	118	50	22
78	1651	856	541	768	882	190	141	135	41	44	130	122	55	27
79	1764	898	580	826	978	202	143	132	38	45	132	122	51	29
80	1650	830	530	768	911	191	147	145	46	47	126	114	48	22
81	1666	837	544	746	924	193	142	137	37	47	127	120	47	26
82	1636	813	521	781	917	199	146	135	33	43	117	121	48	22
83	1671	815	543	789	944	207	149	150	40	49	122	118	50	20
84	1793	892	584	874	1019	189	139	135	38	46	132	121	49	34
85	1706	801	502	822	970	207	141	136	39	52	129	121	49	25
86	1649	820	531	758	908	189	142	136	37	42	120	123	49	29
87	1693	831	544	833	973	190	143	135	37	45	122	121	49	29
88	1662	873	577	769	897	201	141	130	36	42	123	119	52	21
89	1724	860	539	806	962	209	154	147	40	48	127	130	55	32
90	1677	800	510	834	943	191	145	132	38	48	121	123	55	32
91	1650	833	538	738	911	194	145	138	30	44	131	123	50	25
92	1675	794	500	780	946	189	141	137	43	38	127	113	48	21
93	1553	798	506	722	820	193	140	137	34	44	120	112	48	30
94	1685	849	562	784	1027	196	142	137	36	42	117	122	50	23
95	1575	782	485	742	868	196	138	125	36	44	130	113	49	20
96	1599	809	524	769	890	192	140	131	34	45	121	119	50	27
97	1549	787	498	712	842	194	149	138	38	48	125	105	41	25
98	1617	798	515	795	921	189	141	137	40	44	118	121	47	22
99	1723	884	558	771	1015	197	138	138	39	45	116	121	50	30
100	1677	845	554	798	929	197	144	136	37	46	119	117	54	28

RACIAL GROUP : BANTOMORPH (*continued*)

BAGANDA (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
101	1653	812	517	786	894	195	142	140	38	43	125	107	45	32
102	1525	794	510	724	815	181	146	136	30	44	121	120	55	21
103	1713	852	546	805	930	201	138	138	34	45	124	123	48	30
104	1625	801	512	757	865	202	133	133	34	43	117	112	51	25
105	1706	842	542	803	951	190	135	132	35	45	120	116	49	23
106	1524	761	477	739	826	191	136	125	32	40	118	124	52	28
107	1664	829	529	806	912	200	151	145	38	49	124	113	47	19
108	1685	814	533	801	952	198	152	143	38	45	124	119	47	23
109	1673	819	516	768	923	199	147	145	43	50	130	120	48	26
110	1644	795	500	784	932	199	144	137	36	49	132	126	52	24
111	1750	849	527	849	995	208	144	138	40	44	131	125	50	27
112	1748	848	548	844	998	190	147	141	37	46	126	113	47	28
113	1689	842	556	824	931	194	137	128	37	43	123	127	48	32
114	1586	843	529	708	814	196	137	132	37	45	114	113	46	23
115	1663	822	519	775	891	196	143	137	32	48	117	119	46	25
116	1777	892	572	854	977	201	144	138	42	48	128	125	49	30
117	1640	792	514	807	909	198	145	134	38	45	117	108	47	24
118	1701	829	523	815	923	196	142	138	32	41	132	114	51	31
119	1665	840	548	805	925	195	152	137	36	43	127	127	51	26
120	1671	852	560	754	901	194	145	137	40	46	125	118	52	20
121	1710	844	540	819	952	206	142	142	37	45	119	114	48	28
122	1636	850	553	764	885	195	143	132	35	45	131	120	51	26
123	1553	807	523	750	832	193	145	130	35	45	118	116	50	27
124	1706	887	579	811	936	193	141	127	35	43	129	114	48	22
125	1669	861	567	818	941	195	135	134	30	44	121	118	50	23
126	1495	779	511	722	789	187	140	133	36	42	119	108	48	26
127	1758	841	546	843	1002	199	150	141	36	45	125	126	56	21
128	1653	876	582	759	884	196	141	133	36	46	110	112	47	28
129	1742	857	552	785	954	200	138	133	39	47	113	132	52	24
130	1609	813	532	746	882	186	147	136	35	42	106	111	49	26
131	1779	864	558	842	993	195	149	145	42	47	120	118	48	27
132	1659	832	532	770	918	188	148	139	37	41	117	117	46	24
133	1617	855	568	747	855	189	148	135	38	44	111	111	48	22
134	1746	871	578	807	953	208	146	141	38	45	128	116	49	24
135	1585	792	529	752	876	182	133	132	37	42	113	114	47	21
136	1621	819	534	776	911	185	139	130	29	47	131	102	48	25
137	1687	856	559	790	931	196	140	133	35	43	121	127	54	26
138	1626	840	538	722	873	196	146	138	38	41	117	131	46	26
139	1644	846	536	759	893	198	148	138	34	45	119	114	47	23
140	1650	833	550	732	884	190	142	137	36	43	118	119	49	21
141	1675	837	532	772	935	211	141	138	36	50	121	116	48	25
142	1699	820	529	792	965	185	141	140	38	43	112	122	55	23
143	1605	852	550	708	852	188	142	136	36	40	122	120	49	24
144	1623	852	553	738	856	189	136	133	38	47	116	112	47	25
145	1615	862	570	721	872	196	132	136	35	43	121	114	52	23
146	1662	839	543	782	903	199	144	139	36	51	122	115	45	25
147	1621	845	543	731	875	208	153	136	36	47	129	119	49	27
148	1752	907	597	803	950	192	144	136	38	46	128	131	54	30
149	1617	820	530	731	883	190	140	128	32	41	112	113	50	23
150	1671	864	565	786	907	196	137	128	39	43	121	124	54	26
151	1729	855	546	777	967	189	148	134	33	43	116	124	54	29
152	1653	860	572	792	882	195	145	137	37	47	118	119	52	31
153	1615	825	536	773	888	197	147	140	34	49	126	113	50	29
154	1692	888	579	725	919	195	142	137	39	44	123	120	51	28
155	1763	869	554	810	966	198	133	139	39	45	116	123	50	25

RACIAL GROUP : BANTOMORPH (*continued*)

BAGANDA (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
156	1603	813	528	716	882	198	136	126	36	45	121	114	50	28
157	1635	830	544	750	903	196	146	136	40	43	122	113	46	26
158	1657	868	580	762	886	199	140	136	34	43	125	118	51	26
159	1623	861	579	776	865	187	139	131	33	42	115	123	53	24
160	1686	881	592	764	914	189	146	143	37	46	128	127	52	26
161	1714	890	584	826	946	202	143	134	36	48	115	131	49	27
162	1652	878	579	788	886	200	140	134	35	45	117	116	46	26
163	1621	828	527	736	899	190	135	132	38	47	117	116	48	26
164	1727	903	581	767	930	200	143	133	40	46	121	127	53	27
165	1606	888	590	728	862	193	141	130	36	41	125	123	51	22
166	1669	886	572	767	901	194	139	130	34	43	120	125	51	31
167	1725	863	563	806	969	192	148	148	37	46	121	124	48	26
168	1720	903	609	770	937	198	137	138	39	47	116	125	55	27
169	1670	860	555	767	915	198	146	144	39	41	131	126	49	28
170	1595	866	569	748	841	190	140	137	37	42	116	115	53	26
171	1616	873	584	716	852	187	145	132	37	46	125	117	52	26
172	1674	863	578	752	916	184	147	138	35	42	126	111	52	22
173	1695	860	559	805	940	206	146	141	40	51	121	123	52	28
174	1732	896	600	796	957	209	141	142	38	46	123	123	51	28
175	1598	792	508	714	876	190	136	125	36	43	117	117	49	26
176	1545	798	491	713	819	191	142	132	36	43	122	118	50	26
177	1636	831	531	791	919	199	145	132	37	43	132	125	49	26
178	1682	854	559	777	922	202	143	136	37	45	117	117	51	26
179	1612	832	537	726	868	200	141	131	35	44	121	117	44	30
180	1527	819	535	696	811	186	134	120	30	43	128	113	48	21
181	1583	814	531	717	846	199	137	127	34	43	121	123	47	29
182	1710	880	592	808	918	190	135	135	35	47	116	116	50	25
183	1655	868	580	778	866	187	138	138	34	45	112	120	52	24
184	1734	896	599	816	957	192	150	144	34	47	115	119	50	26
185	1611	842	555	789	886	185	132	127	33	43	115	117	48	31
186	1684	827	543	807	969	192	144	142	40	51	115	122	54	24
187	1645	839	534	826	904	195	141	138	34	41	114	118	49	21
188	1732	874	570	851	966	192	135	131	40	47	123	114	47	30
189	1688	850	560	819	942	202	140	128	34	45	127	121	52	26
190	1598	807	528	732	887	197	136	126	37	44	120	113	48	28
191	1649	865	566	782	896	196	145	140	41	46	121	114	47	24
192	1601	855	556	727	824	198	136	136	44	47	126	121	49	20
193	1587	817	519	760	853	192	150	137	41	46	127	121	48	20
194	1707	842	543	861	976	202	145	136	36	45	124	116	50	26
195	1711	884	558	782	912	205	150	141	37	41	126	121	48	27
196	1806	887	590	879	1043	211	146	144	40	49	124	119	49	22
197	1716	870	562	788	947	200	149	143	43	43	128	114	53	28
198	1611	831	539	789	884	192	137	136	35	45	120	123	49	24
199	1645	848	571	770	900	195	144	138	35	44	130	122	50	29
200	1676	884	600	774	903	207	146	140	33	47	116	133	55	24
201	1648	854	541	758	900	195	148	143	32	46	120	115	46	26
202	1655	831	514	798	922	188	137	135	37	48	111	115	51	23
203	1647	846	549	770	901	196	145	141	35	45	118	116	48	23
204	1690	848	549	740	917	202	137	131	35	46	120	117	49	33
205	1501	796	514	703	804	198	137	141	32	42	127	123	53	34
206	1646	842	553	784	910	196	146	144	39	44	116	121	47	24
207	1675	850	553	775	922	203	150	148	41	47	109	123	56	28
208	1605	852	559	756	864	196	145	134	33	45	127	115	47	27
209	1770	911	609	816	975	193	141	135	39	48	124	117	54	29
210	1771	916	580	825	957	196	142	140	35	44	132	122	49	29

RACIAL GROUP : BANTOMORPH (*continued*)

BAGANDA (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
211	1482	796	529	664	760	190	137	130	36	40	119	104	48	26
212	1593	838	528	721	834	187	143	132	32	36	119	111	50	19
213	1680	878	583	811	929	187	145	134	35	47	117	125	52	26
214	1645	868	547	809	899	202	145	147	36	47	133	131	53	31
215	1596	809	539	779	870	195	150	145	37	46	117	116	48	31
216	1659	845	542	762	916	190	144	130	33	46	123	115	46	26
217	1694	877	568	792	867	196	147	145	37	49	132	130	52	32
218	1673	877	576	783	896	192	140	135	34	44	114	117	47	23
219	1639	843	555	813	913	195	141	138	35	39	118	109	47	21
220	1658	836	536	772	929	188	136	136	33	41	114	117	48	24
221	1729	934	614	787	920	196	135	131	36	46	131	116	51	23
222	1723	852	546	841	958	195	145	140	37	49	119	113	48	32
223	1540	821	531	702	811	198	145	132	31	50	128	115	49	25
224	1623	839	545	780	879	193	146	139	36	48	125	125	51	36
225	1604	835	543	770	876	195	147	133	35	44	130	112	47	28
226	1650	822	525	784	932	181	140	139	34	50	119	119	47	31
227	1618	826	538	789	872	200	136	134	38	45	113	111	47	17
228	1656	876	569	780	909	197	142	133	36	46	109	112	47	21
229	1644	828	541	801	917	195	142	138	39	44	119	114	51	31
230	1605	856	567	769	846	202	142	139	34	46	122	114	47	17
231	1631	859	567	756	886	194	141	137	33	49	126	112	47	25
232	1651	852	567	801	891	195	140	141	37	45	119	110	46	29
233	1691	891	585	801	938	202	145	133	41	49	122	120	48	27
234	1618	833	529	752	876	196	143	138	37	40	129	132	47	27
235	1573	800	517	785	881	182	138	138	34	42	117	111	45	26
236	1607	836	542	757	868	201	142	132	36	46	120	114	47	28
237	1635	853	563	808	881	196	145	135	36	41	120	124	51	28
238	1740	900	589	829	948	195	151	151	43	48	126	122	48	22
239	1608	817	513	758	872	198	140	136	35	44	123	122	49	24
240	1683	850	539	809	913	200	137	133	36	44	114	118	51	25
241	1763	897	592	820	976	195	143	135	39	47	126	120	53	30
242	1564	788	502	786	877	182	138	137	32	41	118	114	47	26
243	1690	865	558	779	941	201	145	132	40	47	120	122	49	28
244	1551	820	536	726	817	188	140	130	37	44	110	122	51	30
245	1663	838	550	792	896	200	145	133	36	47	121	115	49	25
246	1651	835	551	760	893	190	142	136	34	44	122	121	51	20
247	1575	815	528	774	874	183	138	130	32	42	118	113	46	22
248	1646	848	571	786	894	195	145	138	35	46	129	122	50	28
249	1659	863	561	786	891	198	145	139	32	51	114	122	49	28
250	1594	805	526	749	889	192	134	134	38	47	117	108	48	26
251	1740	895	585	826	942	201	141	133	33	47	134	110	49	21
252	1679	875	580	783	901	188	145	140	36	45	128	118	48	32
253	1763	899	599	825	942	195	140	137	36	45	125	117	49	26
254	1575	823	542	706	841	192	144	137	39	42	119	116	50	30
255	1669	844	555	808	936	193	149	141	36	49	128	112	49	20
256	1559	825	527	758	861	204	150	138	38	48	128	127	55	28
257	1696	880	581	799	917	189	143	142	39	47	119	114	48	29
258	1590	842	552	765	867	198	139	135	35	45	121	121	50	24
259	1682	864	560	761	889	201	140	131	37	46	126	116	54	33
260	1748	905	595	828	943	196	151	151	44	46	130	120	46	24
261	1722	874	581	850	948	192	134	132	39	46	124	115	46	33
262	1613	833	538	812	869	193	138	137	34	45	122	121	50	23
263	1557	805	519	774	864	191	134	128	35	42	120	114	47	32
264	1679	875	577	810	918	196	144	138	35	47	119	124	55	20
265	1689	863	558	793	930	200	145	142	38	50	122	121	56	29

RACIAL GROUP : BANTOMORPH (*continued*)

Baganda (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
266	1669	870	564	795	908	190	141	132	32	41	133	130	55	25
267	1670	877	579	773	900	189	144	136	35	47	119	119	56	22
268	1684	858	561	760	934	198	141	140	39	45	121	120	46	29
269	1678	877	588	811	913	198	142	138	34	47	120	122	51	21
270	1784	900	597	866	1003	212	147	139	41	45	127	121	49	26
271	1662	859	546	796	900	205	152	147	38	45	126	116	47	32
272	1682	884	577	804	911	201	142	140	38	45	124	122	46	23
273	1540	769	492	741	858	187	144	138	36	51	119	119	46	15
274	1736	875	563	790	965	198	146	142	41	47	121	131	50	22
275	1726	901	605	809	954	202	146	135	33	46	119	117	49	21
276	1598	784	500	765	922	180	136	125	36	40	116	108	48	28
277	1707	876	566	809	923	202	149	140	38	43	133	113	49	23
278	1645	842	534	779	894	193	149	137	36	39	121	127	50	24
279	1579	836	550	716	826	191	150	140	39	41	125	119	47	25
280	1640	858	568	759	855	203	146	135	36	44	129	122	54	34
281	1621	815	529	783	900	199	141	136	38	46	117	119	50	26
282	1624	803	531	774	903	204	144	138	37	44	123	116	46	33
283	1644	850	541	770	892	200	142	140	36	45	129	130	50	28
284	1633	839	530	770	898	187	138	135	32	43	126	113	49	25
285	1603	890	597	773	882	198	146	132	38	43	120	116	53	28
286	1693	826	528	813	959	195	141	133	31	41	119	117	49	27
287	1674	847	558	803	936	205	138	133	33	43	119	107	51	22
288	1683	863	569	798	939	200	145	141	37	49	124	121	53	30
289	1576	800	513	752	854	191	147	140	37	43	114	118	48	24
290	1635	852	565	774	878	195	141	127	36	42	112	116	47	30
291	1493	785	506	696	793	185	134	127	35	42	114	116	49	30
292	1691	881	575	809	916	201	143	140	37	46	127	125	50	24
293	1642	859	561	773	893	195	141	137	35	45	119	112	48	25
294	1647	858	569	763	876	190	138	135	35	44	129	117	48	20
295	1590	816	530	759	900	192	144	124	35	42	117	114	48	22
296	1651	857	554	801	898	194	144	139	35	43	125	116	49	26
297	1693	868	563	809	930	193	134	132	35	43	128	121	51	26
298	1678	839	530	831	931	195	136	132	33	51	127	119	51	24
299	1700	915	610	786	913	202	147	140	32	46	133	123	48	25
300	1767	913	599	835	965	204	144	148	41	46	130	113	47	27
301	1571	827	541	752	858	194	141	132	35	45	127	127	52	33

Bahutu

1	1682	848	557	811	958	191	148	132	40	46	119	120	52	30
2	1679	853	562	812	945	186	138	140	39	42	117	118	46	27
3	1695	850	547	820	927	197	139	140	38	52	122	115	51	23
4	1679	885	585	779	898	192	141	128	33	42	118	119	47	22
5	1658	853	569	823	945	194	145	140	35	42	125	125	50	27
6	1653	836	551	774	885	201	146	137	38	46	123	124	54	29
7	1634	828	532	772	899	209	147	142	40	44	134	112	49	27
8	1670	856	559	784	907	197	145	141	37	45	121	117	46	25
9	1601	830	549	751	877	184	143	136	36	38	111	108	48	15
10	1691	863	577	811	919	196	144	143	41	47	117	131	50	32
11	1712	905	607	774	924	200	146	135	39	46	118	113	46	26
12	1618	822	541	784	849	181	148	130	29	40	122	111	46	24
13	1613	827	537	754	879	192	147	130	34	44	130	113	48	25
14	1668	848	543	785	916	190	139	137	36	44	119	118	48	32
15	1633	843	557	774	884	190	140	126	34	48	118	126	55	26

RACIAL GROUP : BANTOMORPH (*continued*)

Bahutu (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
16	1710	874	576	819	975	183	145	140	36	44	119	114	52	27
17	1520	792	534	720	817	181	136	125	33	40	105	118	53	18
18	1547	810	532	736	803	187	140	132	36	42	120	111	44	24
19	1712	874	567	791	946	203	146	139	37	45	118	113	47	34
20	1648	838	555	772	929	190	143	137	39	46	112	109	49	24
21	1628	842	566	743	889	190	142	131	39	45	112	110	49	32
22	1645	849	574	756	892	187	146	134	36	45	110	119	49	27
23	1506	786	511	702	837	195	142	134	40	46	116	112	48	26
24	1630	858	564	742	869	190	139	131	38	39	123	115	51	30
25	1638	832	540	746	897	208	149	143	40	45	130	112	46	28
26	1637	838	557	798	904	184	134	133	34	44	115	109	50	24
27	1586	776	504	766	903	193	140	127	35	43	118	111	46	26
28	1671	865	574	803	939	189	147	136	35	49	120	121	49	26
29	1680	887	599	784	932	195	147	134	39	48	121	118	48	27
30	1559	822	548	727	830	191	137	129	35	38	112	107	48	26
31	1652	847	554	789	916	198	142	137	37	45	123	113	47	27
32	1658	841	558	769	925	191	143	131	32	41	120	117	48	26
33	1584	844	560	745	859	190	142	132	34	49	120	108	47	20
34	1575	791	519	760	864	187	141	127	38	39	116	111	51	30
35	1649	847	552	762	888	187	141	132	37	43	118	123	48	23
36	1594	801	524	737	895	186	144	130	32	40	122	125	47	24
37	1526	772	512	699	824	179	136	124	33	40	114	111	51	20
38	1694	811	520	848	987	188	142	135	36	44	114	109	44	25
39	1724	843	543	842	980	200	144	136	35	40	120	125	51	30
40	1826	887	584	861	1077	193	144	135	37	40	123	117	53	25
41	1676	862	571	792	922	194	146	141	34	42	119	114	52	26
42	1489	720	465	757	858	186	152	148	37	45	117	116	47	25
43	1751	893	578	847	975	194	148	140	31	45	120	112	52	30
44	1649	827	538	795	903	186	149	140	37	43	116	121	50	27
45	1637	844	576	778	881	193	142	130	30	43	108	112	48	25
46	1706	857	561	823	944	191	144	135	39	43	117	122	48	35
47	1614	815	542	746	887	185	144	131	33	39	114	107	46	24
48	1516	803	525	708	810	193	137	133	33	40	118	113	48	25
49	1579	826	547	766	822	198	143	136	41	44	117	112	47	27
50	1618	876	580	758	872	194	150	136	38	47	124	120	54	29
51	1706	893	598	795	920	190	147	137	37	45	121	116	51	28
52	1632	830	542	781	907	173	135	136	35	44	110	115	47	24
53	1616	815	540	796	883	185	150	136	32	45	111	122	55	31
54	1587	826	548	754	869	186	141	133	39	39	113	137	46	19
55	1759	894	592	807	965	195	146	131	37	40	121	117	52	23
56	1628	825	526	768	911	191	136	134	37	42	121	115	46	29
57	1677	841	557	807	947	191	146	139	35	45	117	114	52	22
58	1585	805	536	788	877	191	140	132	34	41	105	117	49	30
59	1635	824	543	791	915	185	133	133	34	43	110	105	48	26
60	1622	818	536	758	862	184	141	126	34	37	113	115	50	28
61	1586	837	553	750	857	188	145	127	37	45	113	106	47	27
62	1702	872	576	779	939	185	147	130	32	41	118	109	48	27
63	1681	847	557	822	958	183	137	132	36	49	124	119	53	35
64	1601	806	520	757	901	186	138	133	35	39	117	102	42	29
65	1660	840	551	787	926	192	147	136	34	42	113	116	49	22
66	1702	850	552	821	934	186	138	134	36	41	123	116	46	27
67	1698	900	601	818	918	202	145	136	37	47	122	118	47	25
68	1701	838	529	864	953	182	140	138	34	43	117	121	50	22
69	1564	814	522	751	844	192	146	141	38	41	123	115	51	26
70	1670	829	544	829	942	191	148	141	35	40	112	133	54	26

RACIAL GROUP: BANTOMORPH (*continued*)

BAHUTU (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
71	1708	863	568	769	963	191	140	142	38	42	123	123	48	28
72	1522	779	492	692	818	192	133	135	30	41	122	121	51	23
73	1714	863	562	815	968	191	139	134	35	44	116	127	48	24
74	1618	827	542	776	897	189	147	141	39	46	125	121	49	26
75	1678	852	558	795	936	192	139	134	36	40	115	111	52	26
76	1673	881	580	806	948	189	143	141	35	47	105	129	50	24
77	1788	934	605	845	972	192	142	130	37	38	117	132	55	24
78	1549	796	523	796	846	182	137	130	39	43	119	110	47	28
79	1647	883	592	768	911	198	142	136	35	44	121	112	45	28
80	1674	873	579	749	892	197	145	140	44	46	116	123	49	27
81	1651	824	533	775	918	182	137	131	34	40	114	102	49	23
82	1643	831	554	740	890	189	136	130	35	42	121	110	47	30
83	1629	867	577	781	888	187	142	131	39	43	117	119	47	24
84	1640	860	575	765	877	194	144	136	33	46	114	113	51	22
85	1695	879	576	765	918	191	146	129	32	39	120	115	49	27
86	1673	889	585	793	918	195	151	142	34	45	113	124	53	28
87	1629	798	502	788	920	180	140	133	40	47	122	116	52	28
88	1552	844	571	682	803	191	149	136	34	46	124	115	49	25
89	1576	878	579	748	827	196	148	135	38	48	125	114	43	30
90	1558	844	557	696	810	195	145	140	34	44	114	116	47	26
91	1552	844	562	701	831	194	143	134	34	44	114	116	48	27
92	1688	900	604	773	899	199	145	135	35	45	118	121	53	27
93	1610	865	581	753	866	196	141	130	36	39	115	120	48	25
94	1645	848	558	765	910	189	141	131	35	43	123	103	45	25
95	1610	846	564	777	883	190	142	136	39	43	102	113	46	26
96	1637	825	538	779	908	193	146	144	35	47	115	118	50	28
97	1705	833	539	855	951	188	143	130	29	42	108	116	49	27
98	1666	841	549	825	898	186	141	140	33	48	112	132	56	30

BATORO

1	1679	860	618	795	930	199	142	136	41	48	122	114	48	26
2	1825	914	596	879	1023	198	148	144	38	44	121	117	53	27
3	1658	846	544	788	925	200	144	134	38	40	121	124	47	27
4	1640	857	582	783	883	194	145	140	38	45	114	115	51	26
5	1732	904	587	791	929	197	136	126	33	44	121	119	53	29
6	1649	825	524	797	918	198	140	140	34	48	130	118	50	23
7	1679	822	537	796	937	194	148	141	35	46	131	121	50	28
8	1731	903	600	839	913	197	151	142	36	41	116	115	52	19
9	1676	850	550	793	928	199	141	135	41	47	124	115	54	25
10	1689	826	545	825	973	191	153	144	35	43	118	124	49	29
11	1673	860	555	763	926	192	147	136	39	45	119	109	48	23
12	1771	904	597	820	971	195	142	140	33	46	125	121	48	28
13	1531	813	538	731	822	193	152	136	34	42	122	123	54	31
14	1673	856	561	798	927	195	140	137	36	42	119	125	55	28
15	1674	855	558	791	928	195	140	137	35	43	126	126	54	26

RACIAL GROUP: BANTOMORPH (*continued*)

Batoro (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
16	1723	846	553	851	987	195	140	131	29	42	117	121	51	28
17	1676	847	553	805	929	199	142	136	42	48	124	116	54	24
18	1747	878	576	819	949	201	144	135	36	41	125	119	52	29
19	1604	826	539	739	872	184	143	140	35	44	119	116	52	25
20	1643	842	539	780	912	190	145	141	33	45	126	117	51	21
21	1653	852	559	756	911	192	145	138	40	44	120	133	53	30
22	1663	840	554	797	921	195	153	141	38	43	115	125	52	28
23	1642	820	524	754	902	197	145	140	41	47	130	112	49	29
24	1625	843	551	756	883	180	144	138	36	39	116	113	50	24
25	1757	916	607	803	950	190	143	135	37	45	119	119	54	26
26	1638	842	545	780	858	181	139	134	34	43	112	122	48	26
27	1542	821	548	741	845	193	148	147	33	44	120	119	51	26
28	1631	815	521	727	912	191	135	132	35	41	110	124	55	22
29	1663	870	564	758	903	193	148	142	40	46	119	117	54	32
30	1650	811	526	769	928	200	141	137	39	45	108	118	51	31
31	1734	856	564	815	984	179	146	138	34	42	123	113	51	25
32	1652	880	576	775	900	194	144	128	31	44	110	127	51	25
33	1640	825	540	769	860	185	137	131	35	44	117	114	52	23
34	1621	844	549	775	880	186	147	135	35	51	112	112	52	21
35	1568	808	522	763	830	186	138	131	32	45	120	111	49	26
36	1828	886	587	767	882	184	140	135	34	45	126	114	49	26
37	1727	870	572	837	961	194	145	141	36	41	117	115	48	30
38	1850	917	600	848	1013	193	142	138	37	44	128	114	50	29
39	1647	817	521	785	920	191	143	137	37	46	119	131	57	28
40	1655	835	540	804	915	192	146	136	33	43	116	113	50	23
41	1740	912	585	800	926	192	143	132	35	44	120	125	53	37
42	1615	870	598	782	885	185	145	134	34	46	114	112	51	27
43	1640	852	531	717	878	195	142	127	38	44	124	118	49	18
44	1676	874	581	777	917	190	146	140	41	44	123	118	49	29
45	1619	857	566	745	877	202	144	138	40	42	132	117	50	24
46	1638	819	519	781	897	201	142	136	35	42	121	116	50	24
47	1715	858	565	796	966	183	136	135	39	41	115	128	56	30
48	1627	824	531	790	913	181	145	143	40	42	121	116	53	24
49	1671	875	577	779	902	196	147	141	43	46	125	118	48	29
50	1746	892	586	795	966	193	145	140	43	42	121	125	47	28
51	1777	874	570	855	1048	198	140	136	36	45	122	112	52	27
52	1759	900	600	843	982	195	148	145	38	46	113	122	51	28
53	1612	825	555	747	899	187	138	127	29	44	108	118	50	27
54	1747	906	593	779	949	196	141	131	33	38	123	120	53	25
55	1635	834	543	710	884	187	145	144	35	43	116	122	52	25
56	1623	823	550	808	901	198	146	139	37	43	125	126	49	25
57	1712	876	595	784	965	187	138	137	33	42	112	117	53	32

Bahiru

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
1	1635	768	471	774	923	193	137	127	35	45	114	127	47	24
2	1748	888	597	860	946	201	140	141	40	45	120	117	50	28
3	1543	813	516	697	835	179	135	130	38	49	119	118	45	21
4	1548	837	558	722	819	183	137	128	35	43	118	119	49	24
5	1628	835	553	747	871	187	141	135	32	42	112	107	46	26

RACIAL GROUP: BANTOMORPH (*continued*)

Bahiru (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
6	1596	795	523	766	877	192	135	129	36	46	115	114	48	24
7	1562	798	510	744	873	186	136	126	35	41	119	114	51	28
8	1610	817	532	737	877	185	141	135	34	43	111	106	48	23
9	1738	884	579	801	975	188	139	130	37	47	120	122	51	31
10	1544	798	527	725	834	185	134	122	37	39	109	109	46	26
11	1636	814	529	774	919	190	137	135	35	40	111	117	46	25
12	1712	846	548	792	943	202	141	133	37	50	118	121	54	21
13	1649	861	552	748	865	192	135	134	38	46	123	123	47	26
14	1637	820	534	759	885	193	141	137	35	39	116	121	47	29
15	1732	856	552	831	959	191	142	139	36	47	116	120	55	19
16	1650	825	549	781	928	196	142	131	40	44	125	121	51	25
17	1750	866	566	839	1020	192	152	136	39	45	123	119	47	29
18	1734	887	571	786	941	196	144	139	39	40	121	118	45	28
19	1609	805	515	762	900	196	134	134	34	46	113	107	50	26
20	1697	842	556	790	927	191	132	127	38	42	119	121	52	27
21	1716	901	579	801	923	195	140	135	33	46	134	127	52	27
22	1677	852	563	807	943	190	136	136	34	40	110	117	55	24
23	1691	876	567	761	922	201	145	140	37	46	123	120	47	25
24	1622	818	527	773	894	178	146	136	35	44	109	111	50	25
25	1671	871	567	780	909	197	145	131	34	40	130	115	49	27
26	1602	804	516	772	899	198	135	135	34	46	119	112	59	25
27	1672	873	575	787	904	197	145	132	35	41	128	115	49	26
28	1689	872	572	773	922	187	137	126	32	42	115	126	48	26
29	1648	825	532	789	930	187	141	136	40	38	118	106	47	21
30	1680	852	584	779	927	193	143	132	31	44	118	114	46	27
31	1645	852	570	780	903	193	136	133	42	45	126	115	49	27
32	1680	831	547	808	946	190	136	136	36	40	115	115	52	22
33	1547	790	523	754	857	183	133	127	33	42	116	108	47	24
34	1546	795	521	739	837	200	135	124	35	42	122	117	46	25
35	1701	868	571	833	924	186	140	137	39	44	124	120	48	27
36	1592	827	545	714	861	185	144	137	39	45	115	112	47	27
37	1549	814	521	697	825	179	137	130	38	48	121	112	50	23
38	1637	814	517	807	911	193	138	127	36	44	116	125	50	24
39	1710	866	565	789	939	198	144	138	40	43	127	116	49	29
40	1752	884	594	820	1003	196	146	133	34	42	114	112	46	30
41	1652	820	539	815	935	181	133	129	36	43	108	115	50	26
42	1546	826	546	715	817	183	138	128	37	42	120	118	50	22
43	1614	828	541	725	859	184	137	136	32	42	124	117	47	25
44	1687	871	568	812	880	200	145	132	37	47	129	126	55	19

Banyoro

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
1	1656	821	542	757	923	192	151	125	34	43	123	110	47	25
2	1694	868	538	733	915	196	141	134	34	46	118	116	51	24
3	1631	880	580	775	857	193	140	135	33	44	125	110	48	28
4	1638	847	566	733	891	191	146	140	41	48	121	114	48	24
5	1637	807	515	768	888	202	142	136	33	42	112	116	49	26

RACIAL GROUP : BANTOMORPH *(continued)*

BANYORO *(continued)*

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
6	1640	824	526	786	905	197	140	133	36	46	113	111	51	20
7	1731	875	577	803	957	205	137	131	41	46	119	113	47	32
8	1769	853	549	820	1046	185	136	127	41	43	119	116	51	25
9	1634	855	577	767	884	196	147	140	37	44	115	115	50	27
10	1727	880	584	792	947	200	146	142	36	47	125	129	52	28
11	1667	856	552	777	912	190	149	141	37	46	124	115	50	29
12	1635	840	549	701	903	189	141	126	37	43	120	110	49	25
13	1633	833	532	781	941	187	142	131	34	42	121	109	51	18
14	1636	868	561	761	863	186	142	132	39	45	115	104	50	21
15	1630	817	523	810	917	183	147	144	41	43	118	117	51	24
16	1690	810	519	783	927	192	145	136	30	44	114	112	48	26
17	1622	849	552	769	878	181	147	131	37	41	120	113	50	25
18	1666	857	565	772	913	188	139	130	39	44	116	113	48	25
19	1612	828	549	745	861	185	137	126	38	48	115	122	51	24
20	1661	848	575	744	903	191	146	135	37	42	110	116	50	26
21	1560	788	516	717	851	187	141	127	31	40	111	109	46	23
22	1621	824	547	697	848	192	140	130	37	42	119	106	45	25
23	1670	824	536	755	914	195	147	126	38	42	123	114	45	27
24	1759	892	582	761	956	199	143	135	33	45	115	116	50	30
25	1776	899	581	836	984	201	146	136	40	43	129	126	48	27
26	1742	887	586	816	968	190	137	142	34	43	127	114	46	29
27	1744	846	546	858	1002	190	140	140	38	47	118	123	50	23
28	1732	827	528	838	986	184	145	136	35	43	118	113	45	30
29	1736	868	563	809	943	188	137	133	36	49	130	124	52	26
30	1630	833	535	759	881	194	139	125	39	49	120	107	48	25
31	1566	814	531	725	859	180	150	137	39	44	128	119	51	26
32	1522	759	498	744	828	189	135	132	29	42	116	115	46	15
33	1622	797	530	776	927	191	151	136	30	37	111	114	47	25
34	1535	798	511	697	810	182	134	122	34	40	106	117	49	23
35	1662	825	519	760	876	193	142	133	34	50	126	107	46	24
36	1756	850	555	808	987	206	143	138	40	51	125	124	55	24
37	1771	874	565	811	968	194	143	141	33	47	121	118	48	28
38	1673	820	524	812	922	193	158	137	39	46	127	117	50	29
39	1610	834	526	754	855	192	148	140	36	45	120	124	53	21
40	1655	848	533	758	888	197	143	126	36	44	126	117	48	25
41	1622	812	521	748	900	195	140	131	33	44	121	113	48	30
42	1740	886	568	786	930	194	143	136	37	45	127	114	51	29
43	1640	844	550	751	917	191	145	132	33	44	117	126	54	28

RACIAL GROUP : HAMITOMORPH

BATUTSI

1	1737	905	588	818	966	198	139	131	37	37	121	121	54	25
2	1764	888	578	820	948	203	144	126	35	38	120	126	52	28
3	1635	854	571	749	876	188	141	132	33	36	119	118	52	28
4	1680	843	538	716	947	191	146	131	36	37	122	114	52	22
5	1734	857	560	810	978	193	138	134	38	44	133	109	45	29
6	1761	890	589	807	958	194	145	130	34	40	129	117	50	24
7	1683	822	527	783	942	196	136	131	35	42	120	119	53	25
8	1649	860	566	796	891	202	143	127	34	40	122	123	49	29
9	1710	831	541	808	953	184	133	131	29	44	122	108	51	24
10	1763	851	528	825	1004	187	136	129	36	37	121	112	49	23

RACIAL GROUP: HAMITOMORPH (*continued*)

BATUTSI (*continued*)

Serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
11	1750	867	574	810	979	187	137	130	36	41	128	127	57	26
12	1781	935	619	791	977	191	144	132	32	35	133	125	56	25
13	1765	852	540	837	996	190	144	128	32	34	125	111	51	25
14	1734	877	591	789	941	201	152	134	35	37	125	114	55	20
15	1724	862	557	777	961	191	136	130	34	41	132	122	50	27
16	1723	894	576	810	961	190	142	136	36	38	113	123	55	27
17	1782	881	584	831	996	191	152	139	30	44	121	121	52	27
18	1675	903	612	752	885	181	137	131	33	37	122	120	48	27
19	1718	892	597	804	945	195	147	138	39	41	131	130	52	25
20	1619	834	553	718	886	183	140	132	36	39	113	105	54	17
21	1766	904	593	768	955	199	144	136	37	43	129	120	50	31
22	1702	893	593	805	916	200	146	135	40	47	122	114	51	26
23	1725	871	575	832	951	186	138	133	35	42	119	121	55	29
24	1661	852	553	778	909	193	141	135	32	36	117	112	46	28
25	1730	857	547	805	958	203	154	145	42	38	131	122	51	26
26	1649	831	542	787	918	191	140	129	36	37	117	121	53	31
27	1665	818	528	799	938	185	138	140	37	39	119	112	49	26
28	1611	855	572	780	880	189	141	136	40	41	124	115	47	26
29	1783	874	567	794	1002	189	146	132	32	41	122	123	52	28
30	1679	847	550	793	910	191	137	134	34	39	118	120	52	28
31	1826	898	590	836	1050	193	145	135	34	43	125	122	61	28
32	1839	948	623	845	1011	200	151	141	35	41	126	125	55	25
33	1835	941	614	809	1040	201	145	134	28	36	124	122	58	24
34	1802	849	548	861	1030	200	145	135	34	44	120	120	57	28
35	1784	892	579	794	984	187	145	135	32	40	120	119	52	28
36	1779	889	575	848	976	202	148	137	36	42	124	115	53	28
37	1871	944	618	894	1095	193	149	141	36	38	133	125	54	28
38	1719	836	550	783	975	195	145	137	32	43	119	115	51	25
39	1739	888	588	802	976	187	135	124	31	40	122	121	57	26
40	1670	835	544	788	931	186	142	138	40	42	118	119	52	29
41	1624	795	523	774	910	184	135	133	42	38	121	118	52	31
42	1709	884	592	802	963	188	136	137	38	40	127	115	51	25
43	1777	893	576	841	1027	196	141	122	32	40	130	123	53	29
44	1737	883	577	813	955	185	142	132	35	42	121	126	54	29
45	1646	820	517	781	894	185	133	118	32	39	112	113	49	25
46	1679	804	519	801	980	184	137	126	33	38	127	122	50	21
47	1671	832	522	780	932	189	137	122	28	37	121	113	47	24
48	1760	929	622	774	990	196	141	124	32	37	123	120	56	27
49	1688	806	489	763	970	194	145	130	31	35	128	114	57	24
50	1679	831	520	773	936	194	141	133	33	40	122	120	51	27
51	1680	913	599	769	902	202	140	128	32	35	118	122	51	33
52	1588	874	596	709	839	197	146	123	39	41	121	107	47	27
53	1664	837	543	764	937	196	136	127	35	38	124	115	51	26
54	1716	888	574	802	924	202	146	141	35	44	121	116	53	30
55	1760	855	550	836	1005	201	145	143	36	40	128	111	49	29
56	1696	846	543	783	925	191	145	133	29	37	130	120	48	26
57	1804	876	581	825	1059	199	141	129	33	37	122	119	50	27
58	1755	875	572	817	989	193	138	119	33	36	124	121	56	25
59	1778	903	614	792	1000	192	152	131	32	34	120	120	47	23
60	1729	865	558	771	960	198	144	131	31	39	127	115	55	28

RACIAL GROUP : HAMITOMORPH (*continued*)

Batutsi (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MEH	NH	LH
61	1648	853	563	750	873	195	148	148	34	43	119	123	55	27
62	1699	874	573	788	932	199	145	137	34	38	118	114	52	26
63	1751	851	556	836	978	187	140	134	35	40	123	121	50	27
64	1667	818	521	781	928	203	144	134	36	45	136	122	52	24
65	1679	831	518	774	937	200	144	130	35	36	127	121	51	26
66	1750	849	546	834	1001	186	150	129	32	39	119	123	54	24
67	1814	890	571	839	1043	196	144	136	36	42	131	119	55	23
68	1788	860	555	832	1027	192	149	140	39	43	124	112	46	24
69	1663	847	549	779	899	201	147	143	39	44	125	118	52	31
70	1726	906	585	790	925	190	143	128	30	37	119	118	55	23
71	1743	867	543	799	980	195	147	135	33	44	125	111	53	24
72	1746	880	589	875	988	192	134	126	36	41	116	113	51	27
73	1638	840	538	748	899	192	141	131	36	37	124	123	50	28
74	1724	834	535	848	999	189	141	128	30	38	118	126	54	27
75	1753	846	562	813	1000	193	143	132	29	42	119	117	52	25
76	1688	856	539	774	920	194	150	134	38	43	121	117	49	24
77	1689	856	552	792	930	190	139	129	33	38	120	118	55	26
78	1740	885	571	786	965	190	145	137	39	38	118	115	50	28
79	1711	838	550	831	969	195	153	139	37	39	125	127	53	31
80	1675	870	562	736	898	192	140	133	33	42	127	115	48	25
81	1741	849	552	827	1020	192	140	141	36	40	110	128	57	23
82	1715	851	551	805	968	203	147	129	35	41	126	125	54	29
83	1739	881	558	784	972	190	143	131	31	37	132	115	52	23
84	1712	867	577	778	952	198	143	132	35	40	121	119	51	28
85	1661	825	539	777	933	183	145	139	33	40	122	116	52	23
86	1715	860	551	822	957	186	148	138	34	40	120	108	51	18
87	1701	851	553	778	955	188	136	124	32	41	118	112	47	25
88	1727	895	587	770	948	190	138	127	34	34	125	114	55	31
89	1768	876	560	798	971	196	146	133	33	38	123	118	54	25
90	1766	878	577	846	996	188	137	129	32	34	128	112	46	23
91	1806	867	566	839	1036	206	144	129	32	42	108	112	48	24
92	1697	843	555	765	928	199	146	130	38	40	125	120	50	27
93	1857	923	601	844	1059	198	149	134	29	39	136	125	56	24
94	1714	875	551	777	936	196	143	131	36	38	128	121	56	29
95	1551	791	502	709	820	201	146	135	38	39	119	108	45	26
96	1751	880	563	830	983	192	140	130	32	40	126	120	51	26
97	1668	821	533	817	932	199	147	136	38	43	130	120	52	27
98	1772	924	620	813	967	196	141	130	34	39	116	133	57	29
99	1704	806	528	829	980	186	133	128	35	39	113	125	60	30
100	1835	887	581	887	1052	195	152	143	36	45	129	131	50	25
101	1729	869	579	778	964	187	146	131	34	42	110	111	55	24
102	1771	902	594	809	986	197	146	133	39	39	129	118	47	25
103	1804	879	574	794	1018	192	147	136	34	38	124	116	50	31
104	1637	850	544	762	902	193	142	130	31	38	120	121	50	32

RACIAL GROUP: MULATTOMORPH

SWAHILI

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
1	1728	898	596	781	946	195	149	141	36	45	124	105	52	26
2	1563	786	513	717	864	192	151	127	39	39	122	106	50	27
3	1758	872	572	815	1004	196	150	134	36	46	124	123	56	29
4	1717	864	559	815	960	199	154	138	39	48	121	121	54	22
5	1811	860	561	876	968	197	148	134	35	44	118	122	59	33
6	1655	847	556	777	910	186	146	133	35	41	119	114	51	20
7	1596	801	513	778	868	193	140	140	35	42	128	116	53	25
8	1736	880	580	814	965	190	140	131	32	44	127	110	51	26
9	1592	828	540	737	865	201	148	131	38	44	120	108	52	24
10	1624	836	552	775	908	185	142	135	35	43	109	109	50	22
11	1598	863	558	761	856	188	142	129	37	47	113	115	51	21
12	1559	823	536	705	812	189	140	130	35	44	112	118	50	30
13	1578	844	557	728	841	200	145	138	38	45	112	116	50	24
14	1666	884	582	752	881	196	152	140	36	51	120	116	52	25
15	1540	771	516	732	867	187	143	134	35	43	108	112	50	23
16	1657	887	583	770	865	193	146	137	34	49	124	131	54	30
17	1585	857	566	722	831	190	142	133	33	39	120	123	53	24
18	1613	860	573	754	839	192	153	133	35	40	129	111	47	25
19	1620	872	581	814	849	187	137	128	36	44	114	115	51	21
20	1595	855	583	747	831	188	145	132	36	47	115	113	53	21
21	1699	880	582	776	910	193	155	138	39	46	126	111	51	29
22	1570	840	541	714	829	201	164	139	38	45	132	120	51	33
23	1689	866	568	782	894	190	148	136	34	46	125	113	51	26
24	1713	915	595	786	926	200	143	142	37	46	129	126	59	30
25	1696	823	525	816	952	200	138	142	39	49	113	117	55	22
26	1662	859	563	799	923	188	156	141	38	47	121	117	53	33
27	1690	850	536	788	918	192	160	144	38	47	130	118	55	23
28	1543	785	508	725	845	198	146	142	38	47	124	115	52	28
29	1689	844	547	812	940	183	147	132	34	41	120	119	51	28
30	1594	820	530	733	833	188	152	133	36	44	125	116	51	25
31	1546	802	522	720	827	181	148	138	35	43	118	116	50	26
32	1614	823	526	735	878	189	155	139	35	43	125	113	50	30
33	1717	865	539	789	943	194	150	140	37	46	121	113	51	26
34	1593	817	536	746	856	197	147	132	33	43	125	124	48	34
35	1660	852	562	755	896	182	140	135	34	43	119	120	55	31
36	1557	781	512	740	859	183	148	127	28	44	124	121	52	20
37	1667	836	533	773	935	193	150	137	34	45	130	116	51	23
38	1749	875	570	835	974	196	151	137	32	41	114	116	52	26
39	1676	823	519	756	911	178	146	133	32	39	130	107	49	28
40	1607	821	509	729	872	185	153	133	35	40	135	106	49	31
41	1609	827	534	775	902	184	142	129	31	48	118	121	51	31
42	1639	802	501	761	922	192	153	130	34	47	124	110	51	26
43	1743	867	530	804	946	204	136	135	38	50	113	125	54	27
44	1645	846	557	754	894	194	141	137	36	42	122	118	48	23
45	1700	861	565	778	952	188	146	134	40	46	127	120	54	23
46	1644	809	514	744	917	193	145	128	35	47	126	110	52	28
47	1663	857	557	754	904	198	152	136	37	41	127	117	54	24
48	1667	839	540	817	907	193	140	125	35	43	126	123	54	30
49	1616	811	524	782	868	197	146	138	36	46	119	113	51	22
50	1625	835	540	752	875	188	137	132	32	39	122	112	49	26

RACIAL GROUP : MULATTOMORPH (*continued*)

Swahili (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
51	1592	833	538	743	864	183	137	134	34	42	121	106	48	26
52	1653	861	554	765	882	188	147	137	36	45	124	116	49	26
53	1650	830	548	762	901	201	148	141	41	45	117	124	54	29
54	1730	845	545	825	982	205	149	145	36	45	118	126	56	28
55	1641	823	525	792	909	187	148	130	45	38	124	115	51	23
56	1582	797	509	730	886	170	145	131	33	44	113	110	51	26
57	1686	859	561	816	929	194	151	139	37	41	122	115	52	28
58	1711	842	567	814	968	194	147	133	38	43	118	123	53	26
59	1755	906	596	824	943	197	152	146	42	40	117	124	59	27
60	1622	854	552	755	864	185	140	130	33	44	118	113	52	24
61	1703	887	566	791	920	193	147	138	34	47	117	116	51	28
62	1698	843	529	773	930	189	145	140	42	46	126	110	50	26
63	1629	822	533	760	892	177	157	138	36	42	121	117	53	27
64	1625	819	523	747	871	185	150	140	33	40	120	112	49	25
65	1593	822	529	720	860	198	143	142	37	43	127	113	48	21
66	1722	865	569	812	954	199	153	137	34	42	121	119	50	27
67	1648	855	555	789	884	202	155	138	37	46	124	119	50	26
68	1659	874	579	755	902	183	156	134	34	43	126	118	48	25
69	1624	852	541	782	866	195	140	133	34	47	130	107	50	30
70	1638	805	519	773	905	199	150	139	38	48	123	107	49	25
71	1645	837	547	771	906	191	148	131	31	41	119	121	55	26
72	1578	800	510	760	848	187	142	133	32	40	123	109	51	26
73	1601	833	546	752	847	189	148	135	41	48	112	119	51	22
74	1615	802	528	743	868	185	140	124	33	40	122	117	52	23
75	1629	822	521	759	876	191	145	145	38	48	122	124	51	30
76	1561	795	514	725	859	190	141	131	31	36	116	114	48	26
77	1601	814	530	764	882	183	141	134	33	43	115	116	49	23
78	1690	830	528	810	963	193	143	134	36	46	127	113	52	31
79	1683	842	551	805	928	204	146	144	39	50	123	120	50	29
80	1577	811	535	734	850	194	148	136	38	45	119	121	50	29
81	1714	835	545	814	958	197	144	133	34	46	122	112	46	23
82	1659	876	570	754	879	184	156	141	37	41	120	118	53	26
83	1698	882	578	792	931	196	152	140	38	44	119	126	52	23
84	1645	837	546	770	905	190	153	131	33	44	122	117	49	24
85	1619	834	538	717	870	183	141	128	41	44	118	122	56	25
86	1718	874	572	833	944	197	151	135	35	45	125	121	50	33
87	1706	886	570	825	938	195	147	139	38	47	130	115	53	27
88	1605	790	500	772	907	183	149	136	37	43	122	117	53	23
89	1641	799	517	714	936	182	141	127	37	44	118	108	50	21
90	1616	840	545	728	874	181	155	139	30	47	121	114	52	28
91	1678	870	560	796	906	191	142	132	36	43	120	116	50	29
92	1814	888	562	902	1007	198	157	148	37	47	121	125	52	26
93	1711	852	543	835	939	210	156	140	36	43	127	131	54	26
94	1668	848	553	764	918	191	153	136	34	40	117	114	57	23
95	1592	821	523	741	861	190	142	123	32	42	119	118	50	29
96	1679	863	566	779	904	199	143	127	35	44	119	122	60	23
97	1644	824	554	769	942	201	146	137	39	39	114	116	53	26
98	1533	775	505	752	860	193	151	129	34	43	117	108	51	24
99	1658	821	525	776	895	182	143	131	33	41	124	114	48	25
100	1678	816	530	841	973	200	149	133	34	49	122	109	53	24
101	1761	897	600	840	942	208	156	151	39	40	141	126	53	20
102	1652	853	551	768	890	197	147	141	36	47	116	116	52	26
103	1627	833	539	749	912	196	145	139	34	47	122	113	53	27
104	1576	838	542	705	838	186	142	135	38	47	116	106	51	26

RACIAL GROUP: NILOTOMORPH

Luo

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
1	1785	900	589	852	997	176	146	131	35	47	116	129	54	30
2	1791	902	579	799	984	208	153	139	40	44	127	119	51	26
3	1815	896	577	819	1029	192	144	142	42	47	126	117	48	29
4	1702	871	584	771	943	186	142	127	37	41	122	107	46	23
5	1667	865	557	762	918	201	150	140	36	47	140	130	49	31
6	1665	819	534	765	920	188	146	144	35	45	119	114	47	25
7	1735	837	532	824	992	204	154	138	42	43	130	123	50	36
8	1631	865	547	740	876	187	151	138	41	46	121	109	49	27
9	1765	900	607	863	962	197	152	138	35	45	125	129	55	32
10	1801	904	597	829	989	197	148	137	33	47	129	108	51	28
11	1794	903	592	819	1002	199	141	134	35	45	126	109	50	22
12	1755	883	570	831	981	199	144	142	43	41	118	126	55	30
13	1756	900	586	783	956	195	149	136	36	47	117	119	50	35
14	1653	846	546	790	904	184	146	140	34	50	119	120	52	21
15	1663	823	529	781	919	188	147	140	43	48	121	111	52	23
16	1824	878	541	837	1033	189	143	137	43	47	122	119	49	31
17	1727	899	591	791	909	190	160	148	37	46	122	115	55	22
18	1725	855	547	761	939	195	150	141	38	43	116	119	53	30
19	1725	867	555	811	938	198	150	137	34	46	121	120	52	29
20	1727	837	504	784	976	200	153	140	39	44	131	123	51	33
21	1672	850	560	795	921	189	146	140	43	47	122	110	51	23
22	1602	776	496	725	872	184	139	131	37	43	117	108	48	26
23	1693	842	548	777	951	195	144	135	35	48	114	115	51	26
24	1817	898	584	821	1006	193	145	137	36	45	121	118	53	24
25	1674	834	516	783	925	191	140	136	34	45	120	120	53	25
26	1720	867	567	789	968	187	141	134	33	43	122	112	52	17
27	1724	848	543	783	951	200	148	136	35	42	120	118	52	18
28	1735	903	584	787	915	205	151	138	34	47	126	117	55	23
29	1852	926	623	866	1046	197	143	143	37	49	116	116	53	28
30	1859	926	594	869	1063	196	147	147	40	48	121	120	51	24
31	1902	930	605	911	1080	198	148	139	36	46	126	133	55	26
32	1702	860	557	774	930	186	155	136	37	46	124	117	49	31
33	1696	836	548	803	937	203	153	141	36	49	121	117	53	26
34	1695	845	542	810	948	191	151	141	35	45	119	114	50	24
35	1676	858	557	754	911	188	137	131	37	48	120	121	58	25
36	1762	893	580	802	955	188	155	135	35	42	120	117	52	25
37	1717	855	538	788	957	191	148	135	37	45	118	130	55	33
38	1730	864	574	853	979	181	148	132	34	43	122	120	51	27
39	1737	861	559	794	974	190	148	137	43	48	124	115	52	30
40	1737	864	551	829	957	200	145	145	40	48	133	127	52	29
41	1737	848	525	772	970	208	152	146	44	47	130	118	52	26
42	1739	838	527	814	966	193	148	137	38	44	125	119	55	24
43	1740	887	564	799	959	194	146	136	36	47	118	118	52	31
44	1737	852	546	806	984	196	149	141	36	44	128	129	51	33
45	1753	892	593	831	942	196	151	142	44	48	121	108	48	24
46	1719	863	568	810	956	197	140	137	37	44	117	111	46	29
47	1744	906	596	826	997	198	138	135	35	43	119	112	51	25
48	1712	879	575	818	944	199	150	141	38	48	130	121	51	21
49	1764	902	588	788	976	187	146	130	35	42	122	112	45	25
50	1764	914	589	818	983	206	149	146	36	47	127	117	50	24

RACIAL GROUP : NILOTOMORPH (*continued*)

Luo (*continued*)

serial no.	St	SH	STH	TUL	TLL	HL	HB	BzB	IoB	NB	HH	MFH	NH	LH
51	1728	843	530	801	974	197	145	135	36	43	124	116	48	27
52	1729	867	552	761	967	185	146	136	41	47	117	117	49	28
53	1704	842	544	806	947	199	153	144	42	47	126	112	50	24
54	1711	885	583	778	935	191	150	135	37	46	124	119	52	29
55	1784	888	582	872	999	192	149	146	39	43	124	124	49	29
56	1802	922	595	797	982	207	152	148	45	50	129	113	50	21
57	1765	887	589	828	982	206	147	144	37	51	127	120	52	29
58	1731	888	564	792	947	204	156	137	37	45	129	121	51	33
59	1731	880	586	820	960	198	141	136	36	42	120	115	49	27
60	1670	810	503	775	911	195	151	135	31	46	125	123	48	21
61	1681	861	542	774	900	201	150	138	31	46	133	128	52	23
62	1690	853	555	808	942	191	136	131	37	42	121	121	50	30
63	1685	837	532	804	930	195	140	139	44	44	120	117	53	28
64	1703	869	559	786	917	191	136	134	41	44	127	121	52	33

MISCELLANEOUS

SCIENCE AND NATIONAL PLANNING*

By P. C. MAHALANOBIS
Indian Statistical Institute, Calcutta

INTRODUCTION

1. In recent years the Anniversary Meeting is being held separately from the Annual Meeting ; and the tradition has been for the President to select a subject for the Anniversary Address from some branch of science of which he is a student and give a general review acceptable to other scientists. As I cannot claim to have specialized knowledge of any particular subject, I looked at the Presidential Addresses of some of my distinguished predecessors for guidance. From the inauguration of the Institute in 1935 I found a continuing concern with problems of organization of scientific research in India. The functions and responsibilities of the National Institute of Sciences have also been reviewed a number of times. Looking back I had the feeling that many of the problems are still with us in one form or another.

2. In 1938, exactly 20 years ago, Meghnad Saha selected for the Presidential Address the "Problem of Indian Rivers," and pleaded for the initiation of systematic studies to lay a scientific foundation for flood control, irrigation, and navigation. He also gave a full discussion of benefits which would accrue by generation of cheap electrical power out of the energy of running water. He had mentioned at that time: "This is a subject which is just beginning in India but its importance has not been properly realized." He referred to poverty and unemployment; and asserted that "industrialization [was] the only solution of the poverty problem." He also referred to planning in Russia, and pointed out that "when the Supreme Council of the Soviet adopted Lenin's resolution for the electrification of the country, it was the USSR Academy of Sciences, and not a committee of bureaucrats which was requested to give a plan." He concluded his address with the following words :

> "If we desire to fight successfully the scourge of poverty and want from which 90 percent of our countrymen are suffering, and lay the foundation of a strong and progressive national life, we must make the fullest use of the power which a knowledge of Nature has given us. We must rebuild our economic system by utilizing the resources of our land, harnessing the energy of our rivers, prospecting for the riches hidden under the bowels of the earth, reclaiming deserts and swamps, conquering the barriers of distance and above all, we must mould anew the nature of man in both individual and social aspects, so that a richer, more harmonious and happier race may live in this great and ancient land of ours."

3. The picture in the mind of the scientist twenty years ago has already become real in the great river valley schemes of India. This is the task of the scientists, to formulate the shape of things to come, not as wishful thinking but based on the

---

*Anniversary Address at the National Institute of Sciences of India by the President, Professor P. C. Mahalanobis on 5 January 1958.

knowledge of Nature and worked out in accordance with the methods of science. The two basic problems, poverty and unemployment, mentioned by Saha still remain as acute as ever. The only way out is through rapid industrial development based on science and planning. I feel this is an appropriate subject for my address.

"Sputnik"

4. On the 4th of October, 1957, an artificial satellite "Sputnik" was launched by USSR followed in a month's time by a second and bigger one carrying a live dog. We received this news when our Council was in session; and on the 6th October, at the desire of the Fellows of the Institute assembled at the Annual Meeting, we sent a message of congratulation to the USSR Academy of Sciences.

5. The way to outer space beyond the earth's atmosphere was thus opened for the first time. It is a great triumph of science. Epochs of history can no longer be demarcated in terms of dynasties or of war between nations but must be thought of in terms of the progress of science and technology.

6. The upsurge of science in what we call the West (that is, in the countries of Europe and North America) led to the industrial revolution characterized by the use of machinery driven by steam, and later by electricity, for the production of goods, and brought about a rapidly increasing level of living in the industrially advanced countries during the last 100 or 150 years. This increase in industrial production gave the West its superiority in armaments; and brought practically the whole world under its domination. Striking developments occurred during the last fifteen years. Atomic energy (for bombs and for electricity), high-speed rockets, guided missiles and satellites have been developed and have great dangers. They also have much of promise for the future, if we do not get involved in a suicidal nuclear war.

7. As scientists, it is very much our concern to understand the significance of recent events. Russia was practically an underdeveloped country at the time of the Revolution forty years ago. The launching of the "Sputnik" from this country became possible only because of the revolutionary progress of science and technology in Russia in the course of one single generation. In Russia, as in all other industrialized countries of the world, scientific research and industrial production have been closely coupled; progress of one has been both cause and effect of an advance of the other. It is this integrated advance of science and industry which has made it possible for USSR to attain its present leading position in the whole world.

Scientific research and economic development

8. In all industrialized countries we find a close association between the attention given to scientific education and research and the rate of increase of industrial production and, hence, of the improvement in the standard of living. Consider the three nuclear Powers, UK, USA and USSR, and also two underdeveloped countries, China and India, which have the two biggest populations in the world and much natural resources. Table 1 shows the population, production of steel, and output of energy in 1955 in these five countries. The equivalent total energy in terms of million tons of coal is given in col. (7) which includes all forms of commercial energy but does not include fuel wood and dung. In Table 2, col. (3) gives the per capita national income at 1955

prices in the currency of the country concerned; col. (4) the per capita production of steel in kilograms; and col. (5) the total commercial energy in equivalent kilowatt hours. Approximate data on expenditure on scientific research are given in cols. (6), (7) and (8). The actual expenditure (expressed in rupees at official rates of exchange) is given in col. (6); and is shown as a percentage of the national income in col. (7). Finally, the expenditure on research per person is shown in col. (8). Sources are mentioned briefly in the footnotes to the Tables; and more detailed information and comments are given in two Statistical Notes (1) and (2) at the end of the Address.

TABLE 1. POPULATION AND PRODUCTION OF STEEL AND ENERGY : 1955

country	population in million	electricity	steel (ingots)	coal	oil	energy (all forms) in terms of coal
(1)	(2)	(3)	(4)	(5)	(6)	(7)
		10^9 kwh	....	..million tons..	....	....
USA	165	629	106.2	448	336	1364
USSR	200	170	45.3	276	71	442
UK	51	94	20.1	225	0.15	249
China	608	12.1	2.9	93	0.79	96
India	382	8.5	1.7	39	0.29	45

Sources: All the figures except population in China are taken from the *UN Statistical Year Book*, 1956 : population, pp. 21–36; steel, pp. 261–262; coal, pp. 138–140; oil, pp. 143–144; electricity. p. 290–294 and energy (all forms), pp. 308–310. The figure for oil in China relates to 1954. The Chinese population figure is taken from statistics prepared by the State Statistical Bureau of the People's Republic of China (in Chinese).

TABLE 2. ECONOMIC DEVELOPMENT AND EXPENDITURE ON RESEARCH : 1955

country	estimates per person per year				approximate expenditure on research		
	national income		steel ingots	energy (all forms)	actual	p.c. of national income	per person
	unit	amount					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
			kg	kwh	Rs. crore	%	Rs.
USA	$	2000	644	13,753	2550	1.8	154
USSR	roubles	5000	226	3,367	2200	1.8	110
UK	£	300	394	8,118	430	2.1	85
China	yuan.	135	4.8	267	65	0.4	1.1
India	Rs.	250	4.5	200	15	0.15	0.38

Sources: Figures in col. (3) are derived from the *Statistics of National Income and Expenditure, UN Statistical Paper, Series H,* No. 10, for USA and UK; *Economic Bulletin of Europe*, May, 1957, for USSR; statistics prepared by the State Statistical Bureau of the People's Republic of China, for China; and *Estimates of National Income*, 1948-49—1955-56, Government of India, for India. Figures in cols. (4) & (5) are derived from Table 1. Figures in column (6) are collected from miscellaneous sources indicated in Statistical Note (2) at the end while figures in cols. (7) & (8) are derived from the figures in col. (6) and those in Table 1.

9. The estimates are admittedly approximate and there are well-known difficulties in international comparisons of income and expenditure. Prices and the structure of prices vary widely from one country to another. Concepts and definitions used in different countries are often different. For example, the American figure for expenditure on research includes industries. It is possible that research carried out in industrial concerns is not included in the Russian figure. Nevertheless some striking contrasts come out clearly in Table 2.

10. The three great powers have the highest per capita income, energy, steel and expenditure on research. China is way behind but has already outstripped India in all items, although only a few years ago China was less developed than India. Expenditure on research is Rs. 65 crore in China compared to Rs. 15 crore in India, or more than four times higher. As a fraction of national income, China spends 0.4 percent which is nearly three times higher than the Indian figure of 0.15 percent. The "per person" expenditure in China is Rs. 1.1 per year which is again three times higher than the Indian expenditure of Rs. 0.38 per year. In 1943 Dr. J. C. Ghosh in his Presidential Address to the National Institute had raised the slogan of "one percent of the national income per year for scientific research." The three advanced countries are spending nearly twice as much at present; and China has almost achieved half the target while, after 15 years, India has gone only one-sixth of the way.

11. Table 3 shows the available number and the out-turn of engineers and doctors in the five countries. Col. (2) gives the population in million to serve as a basis for comparison; cols. (3), (5), (7) and (9) give the years of reference. Col. (4) shows the estimated number of engineers available (in thousands), and col. (6) the annual out-turn (in thousands). In the three advanced countries the out-turn figures refer to training at University level; in China and India the figures include both graduates and diploma-holders. Col. (8) gives the available number of doctors (in thousands), and col. (10) the annual out-turn (in thousands); the out-turn figures refer to graduates in all countries. The corresponding numbers per million of the population are given in the lower half of the table.

12. The three advanced countries have much higher numbers of both "available" and "out-turn" of engineering and medical personnel. USSR has established a definite lead over both USA and UK; the actual number and proportionate out-turn of both engineers and doctors in USSR are more than double those of USA. This indicates a striking superiority in technical man-power, the effect of which is bound to be felt in an increasing measure in future. Scientific and technical man-power requires a longer period of maturing than basic heavy industries. In this respect USSR now occupies, beyond dispute, the leading position in the world. It is reasonable to infer that this lead has some connexion with the launching of the "Sputnik."

13. China and India are much behind the more advanced countries. The available number of engineers and doctors are roughly of the same order in both China and India but China still has appreciably fewer than India on a population basis, namely, 132 engineers per million of the Chinese population against 188 engineers per million of the Indian population with 116 and 183 per million as the corresponding figures for doctors. It must be remembered that engineering and medical education of the modern type had started several decades earlier in India. China, however, has left India behind

in the out-turn of both engineers and doctors. The number of engineers coming out every year is 30.9 per million in China or 60 percent greater than the Indian out-turn of 18.4 per million in 1955. As regards doctors, the out-turn of China is 11.2 per million against 8.1 per million of India, or 40 percent greater.

TABLE 3. NUMBER AND ANNUAL OUT-TURN OF ENGINEERS AND DOCTORS

country	1955 population (10^6)	engineers				doctors			
		total available		annual out-turn		total available		annual out-turn	
		year	number (000)	year	number (000)	year	number (000)	year	number (000)
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
USA	165	1950	518.8	1956	26.0	1955	218.1	1955	7.1
USSR	200	1956	721.0	1956	71.0	1955	299.0	1956	20.0
UK	51	1956	78.5	1954–5	2.8	1956	85.0	1952-3	1.9
China	608	1955	80.2	1955	18.8	1955	70.5	1955	6.8
India	382	1955	71.9	1955	7.0	1955-6	70.0	1954	3.1
		....number of persons per million of population....							
USA			3144		158.0		1322		42.5
USSR			3605		355.0		1495		100.0
UK			1539		55.0		1674		37.3
China			132		30.9		116		11.2
India			188		18.4		183		8.1

Sources: given in Statistical Note (3) at the end of the Address.

14. With the rapid expansion of scientific and technical man-power and of scientific research in USSR there has also been a rapid expansion in the industrial production of the country.[1] The expansion of scientific and technical man-power and of scientific research, on one hand, and rapid increase of industrial production, on the other, are but two aspects of the same transformation. USSR and China (and India) have large natural resources which can be utilized for the benefit of the men who live in these countries. Such utilization or exploitation of natural resources can be done only with the help of scientists, engineers, and technical personnel. It is clear that scientific research and availability of scientific and technical personnel are the important factors in the economic development and "power" of nations, however it may be measured. It is also clear that India is lagging very much behind in these respects.

---

[1] Due to destruction of steel factories during the civil war, the production of steel had fallen to 200,000 tons in 1920 but rose to 18.3 million tons in 1940; dropped to 12.3 million tons in 1945 due to destruction during the war; and rose to 45.3 million tons in 1955 with a target of 68.3 million tons in 1960. Output of black coal (in million tons) increased from 99.4 in 1945 to 276.1 million tons in 1955 and has a target of about 415 million tons in 1960. The output of electricity in billion (thousand million) kilowatt hour rose from 43.3 in 1945 to 170.1 in 1955 with a target of 320 bkwh in 1960. The rate of increase in the USSR is far higher than that achieved in the USA, UK, or West Europe.

In China also the rate of development has been very rapid since the beginning of the First Five Year Plan in 1953. Production of steel (in million tons) rose from 1.8 in 1953 to 5.2 in 1957 with a target of, possibly, 15 million tons in 1962. Production of coal (in million tons) increased from 67 in 1953 to 122 in 1957 and has a target of about 200 million tons in 1962. The output of electricity (in billion kilowatt hour) increased from 9 in 1953 to 19 in 1957 with a target of about 43 in 1962. (All figures have been taken from official Chinese sources).

Economic and social conditions in India

Consumer Expenditure

15. The two great problems of India are poverty and unemployment which are but two aspects of economic stagnation and under-development. Since 1950 the National Sample Survey (NSS) has been collecting information on social and economic conditions in both rural and urban areas of India. Some of this material has been given in the Statistical Appendix from which much interesting and significant information can be obtained. I shall give a few selected examples.

16. The following Table 4 gives in cols. (1) to (4) information on the total consumer expenditure for the whole of India (rural and urban, excluding Jammu & Kashmir) during the period December 1955 to May 1956. Cols. (1) and (3) give the percentage of the population, and cols. (2) and (4) the limiting expenditure in rupees per person per 30 days.

TABLE 4. DISTRIBUTION BY SIZE OF (*a*) CONSUMER EXPENDITURE : 1955-56 AND (*b*) LANDHOLDINGS (1953-54) : ALL-INDIA

	by total consumer expenditure in rupees per person per 30 days[1]				by size of land-holdings in acres[2]			
	percent of population	limit of expenditure in rupees	percent of population	limit of expenditure in rupees	percent of households	limit of size of land-holdings	percent of households	limit of size of land-holdings
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	upto	less than	upto	less than	upto	less than	upto	less than
1.	5 p.c.	Rs. 3.2	55 p.c.	Rs. 15.7	5 p.c.	0.2 acre	55 p.c.	4.0 acre
2.	10 ,,	,, 6.2	60 ,,	,, 16.9	10 ,,	0.5 ,,	60 ,,	4.7 ,,
3.	15 ,,	,, 8.4	65 ,,	,, 18.4	15 ,,	0.7 ,,	65 ,,	5.7 ,,
4.	20 ,,	,, 9.8	70 ,,	,, 20.0	20 ,,	1.0 ,,	70 ,,	6.6 ,,
5.	25 ,,	,, 10.4	75 ,,	,, 22.2	25 ,,	1.3 ,,	75 ,,	7.9 ,,
6.	30 ,,	,, 10.9	80 ,,	,, 24.9	30 ,,	1.7	80 ,,	9.7 ,,
7.	35 ,,	,, 11.6	85 ,,	,, 28.6	35 ,,	2.0 ,,	85 ,,	12.6 ,,
8.	40 ,,	,, 12.4	90 ,,	,, 35.8	40 ,,	2.4 ,,	90 ,,	16.2 ,,
9.	45 ,,	,, 13.4	95 ,,	,, 44.2	45 ,,	2.8 ,,	95 ,,	26.0 ,,
10.	50 ,,	,, 14.6			50 ,,	3.3 ,,		

17. It will be seen from the first line that 5 percent of the population or nearly 20 million spend less than Rs. 3.2 per person per month; and from the second line that 10 percent or more than 38 million people spend Rs. 6.2 per head per month. From the bottom line of col. (2), we find that 50 percent or *half of our countrymen, nearly 200 million in number, live on Rs. 14.6 per month or less than half a rupee per day*[3] *per person.* It would be also seen that even 90 percent of the population spend less than Rs. 35.8 per

---

[1] Based on National Sample Survey (NSS) : 10th Round, December 1955—May 1956, All-India (rural and urban) excluding Jammu & Kashmir.

[2] Based on National Sample Survey (NSS) : 8th Round, July 1954—March 1955, All-India rural area only.

[3] One rupee is about 21 US cents and 1 shilling 6 pence in British currency.

month or just over a rupee a day. That is, *only about 10 or 11 percent of the population of India can spend more than one rupee a day.* This would give some idea of the poverty in which our countrymen have to live.

18. Many other facts regarding conditions of living can be found from the Appendix Tables A.1.0, A.1.1, and A.1.2. The rich actually spend more on food-grains but proportionally less in relation to their total expenditure as can be seen from Table A.1.1. However, both actual and proportional expenditures on sugar, milk and milk-products, and clothing are relatively much higher among the rich, showing that these are essentially luxury items. It is interesting to observe that medical care is apparently the greatest luxury; the poor can have very little of it, and medical expenses go up very steeply as the level of expenditure increases.

Land owned by Households

19. Similar estimates are given in cols. (5) to (8) of the same Table 4 in respect of the size of the holding of lands in acres owned by any assigned percentage of households who own 0.1 acre or more of land. (Households owning less than 0.1 acre were omitted to exclude landless labour and non-agricultural households). It is seen from this table that 20 percent, or one-fifth of the households, have 1 acre or less; fifty percent have less than 3.3 acre; 75 percent have less than 7.9 acre; and 90 percent less than 16.2 acre. Most of the households in the rural area own very little land.

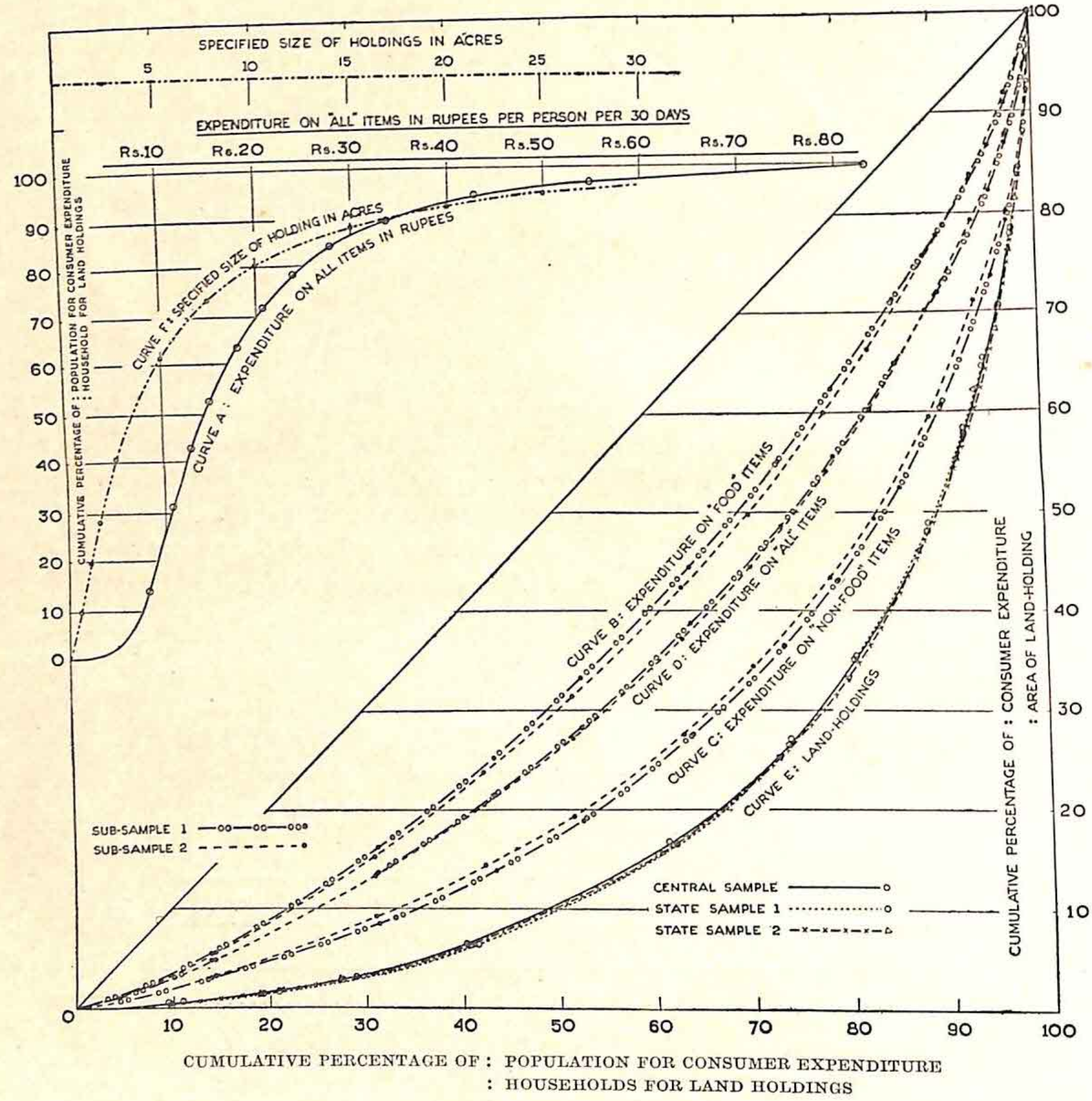
TABLE 5. CONCENTRATION CURVES OF (*a*) CONSUMER EXPENDITURE: 1955-56 : AND (*b*) LAND OWNED BY HOUSEHOLDS 1953-54 : ALL-INDIA

	(a) persons or (b) households	consumer expenditure on all items by persons from		total land owned by households from	
		bottom (0)	top (100%)	bottom (0)	top (100%)
	(1)	(2)	(3)	(4)	(5)
		....................cumulative percentages....................			
1	upto 5 p.c.	1.5	18.2	0.2	34.2
2	10 ,,	3.2	28.8	0.4	48.8
3	15 ,,	5.2	36.6	0.9	58.6
4	20 ,,	7.6	44.0	1.5	66.5
5	25 ,,	10.1	49.9	2.3	72.5
6	30 ,,	12.8	55.3	3.1	77.5
7	35 ,,	15.5	60.5	4.6	81.0
8	40 ,,	18.8	65.6	6.2	84.6
9	45 ,,	22.3	70.0	8.1	87.5
10	50 ,,	26.1	73.9	10.5	89.5

20. The disparity between the rich and the poor can be seen more clearly from the concentration curves shown in the accompanying chart and also from Table 5 above. The percentage of the population or, alternatively, of households is shown in col. (1). In each line, col. (2) shows what percentage of total expenditure (of the whole

CONCENTRATION CURVES

FOR CONSUMER EXPENDITURE AND LAND HOLDINGS



BASED ON NATIONAL SAMPLE SURVEY ALL-INDIA

(Two curves based respectively on two interpenetrating samples 1 and 2 are shown for each concentration curve to indicate the margin of sampling error.)

CURVE A : Distribution by expenditure on 'all' items in rupees CURVE B : Concentration curve for expenditure on 'food' items CURVE C : Concentration curve for expenditure on 'non-food' items CURVE D : Concentration curve for expenditure on 'all' items	10th ROUND (December 1955–May 1956)
CURVE E : Concentration curve for land holdings CURVE F : Distribution by size of land holdings	8th ROUND (July 1954–March 1955)

population) is shared by the percentage of population from the *bottom* as given in col. (1) of the same line; col. (3) gives the share of the same percentage of persons but from the *top*. Thus, the lower 5 percent of the population of India share only 1.5 percent of the total consumer expenditure, while the top 5 percent of the population share 18.2 percent of the total expenditure. We also find that while the lower 25 percent of the population share 10 percent of the total expenditure, the top 25 percent share 50 percent; and that the lower half has 26 percent while the upper half of the population share 74 percent of the total expenditure.

21. Similar estimates are given in cols. (4) and (5) of Table 5 in respect of the proportion of land owned by different percentages of households. The bottom 5 percent own very little land, only about 0.2 percent or less, while the top 5 percent own 34 percent of all lands owned by households. In the same way, it can be seen that the bottom 20 percent have 1.5 percent against 66 percent of all lands owned by the top 20 percent. The lower half of households have 10 percent against 90 percent owned by the upper half. Disparities are much higher at the top. For example, the top one percent of households owns approximately 13 percent of the total land; the top two percent nearly 21 percent; and the top five percent of households own 30 percent or almost one-third of all lands owned by households.[1] The concentration of landholdings is clearly much greater than that of household expenditure.

Employment and Unemployment

22. Some information on employment and unemployment is given in the Statistical Appendix, Tables A.3.1, A.3.2, and A.3.3. I should explain that there are great difficulties in adopting in India the concept of 'unemployment' used in the advanced countries of the world, where practically all gainfully employed persons receive wages and salaries or have earnings in money (except for time spent by members of a family on their own household work). It is, therefore, possible to specify in an objective manner whether a person is gainfully occupied or is out of work. There are, of course, technical difficulties of standardizing the definition of unemployment even in the advanced countries but these are not very serious. In an underdeveloped country like India, on the other hand, there is inevitably a great deal of ambiguity in defining unemployment in an objective manner. Much of the work in household enterprises, in both agriculture and cottage industries, is done by members of the household and is not paid for in money. Also, a large number of persons work on their own account, such as weavers, tailors, carpenters, smiths, barbers, small store-keepers, etc. These people may have enough work to keep them busy the whole day; or they may have to sit idle for a part or most of the time; or may have practically no work in hand; and yet they can never be called technically "unemployed" in accordance with the concept used in the more advanced countries. The above considerations must be kept in mind in interpreting the information given in the Statistical Appendix based on the National Sample Survey covering about 24,700 households in the 9th Round: May–August 1955, and covering about 17,600 households in the 5th and 6th Rounds, December 1952–August 1953. It must also be remembered that all estimates, being based on sample surveys, have margins of errors of sampling.

---

[1] By interpolation from the tables.

23. Subject to these warnings, we may say broadly that about 160 million persons (or roughly 42 percent out of an estimated population of 382 million in 1955) were in the labour force, that is, were either engaged in work or were seeking work. Among them, 2.2 million were "unemployed" in the technical sense of the advanced countries [Table A.3.1], which, however, has no clear meaning in India. In fact, only about 13 million persons, who work under Government and public authorities or in private organized large scale enterprises, have employment of the type usual in the advanced countries [Table A.3.2]. The technical concept of "unemployment" can be strictly used only in the case of these 13 million persons (out of a total labour force of 160 million), but is not applicable to the remaining 147 million who work in household or small scale enterprises.

24. In an underdeveloped country it is, therefore, necessary to consider how many persons have to sit idle a part or most of the time for lack of work and other economic reasons. From the information collected by the NSS it would seem that nearly 20 million persons normally work one hour or less per day, 27 million work two hours or less per day, and nearly 45 million persons are engaged in gainful work for 4 hours a day or less [Table A.3.1, portion B.2]. In NSS, 5th and 6th Rounds: December 1952–March 1953, it was found that nearly 30 million persons have gainful work for less than 5 days in the month, 39 million less than 10 days, and 53 million less than 15 days in the month [Table A.3.1, portion B.3]. In another enquiry, NSS : 7th Round, October 1953–March 1954, it was found that about 45 million persons were working part-time. They were asked supplementary questions; and from the replies received it seems that among them about 10 million persons were not able to work full time partly on account of their own ill-health or on account of the illness of other members of the family; over 23 million (or about half) gave reasons of an economic nature such as lack of demand for their labour, lack of tools and raw materials, slack and off season, etc.; and 13 million gave other reasons. Finally, in NSS : 9th Round, in reply to a specific question, nearly 12 million persons stated that they were seeking additional work and would be available for 4 hours of work or more per day. [Table A.3.1, portion B.1].

25. From the above account it would be seen that it is difficult to give a precise figure for unemployment or underemployment in India. The estimate would depend on which particular concept and definition is used to measure unemployment or underemployment. However, it may, perhaps, be broadly stated that there can be scarcely any doubt that *10 or 12 million persons are either unemployed in the technical sense of the more advanced countries or are severely underemployed. If consideration is extended to those who are sitting idle for more than half or two-thirds of their working time, the number would probably rise to 25 or 30 million persons.*

26. Some more specific information is available regarding unemployment among the educated defined as matriculates and persons having higher qualifications. It is estimated that their total number in 1955 was about 50 lakhs, out of which 4.7 lakhs (or 9.4 percent) were unemployed, most probably in the technical sense of the term as being out of a job or seeking a job as in the more advanced countries. It must be remembered in this connexion that in 1955 just over 4 lakhs of candidates passed the matriculation examination, and the number is rising from year to year. *There is already about 10 percent unemployed among the educated class. The proportion is likely to increase considerably in future which would have grave social and political implications.*

General and Professional Education

27. I shall now briefly consider the question of trained personnel. Some information regarding the number of persons passing the matriculation examination and acquiring other general or professional qualifications every year from 1950 to 1955 is given in Appendix Table A.3.4. Much interesting information can be obtained from this table. I shall make only a few brief observations.

28. There has been fairly rapid all round progress in numbers. The number of matriculates increased roughly from 1.89 lakhs in 1950 to just over 4 lakhs in 1955 or has doubled in 5 years. The rate of increase is about the same at other levels. It is also seen that the number of candidates passing the intermediate examination is roughly 40 percent of the number who passed the matriculation examination two years earlier; the number taking the first degree is roughly 60 percent of the number who passed the intermediate examination two years earlier; and the number taking the master's degree is roughly one-fourth of the number who took the first degree, two years earlier. These are approximate statistical proportions which would vary considerably from year to year but can give some idea of the structure of general education. As a very rough rule, out of 100 matriculates in any given year about 40 would pass the intermediate two years later; 24 would take the first degree after two more years; and 6 would take the master's degree two years later.

29. It is of interest to make some comparisons on the basis of the number of literates in India which was about 63 million in 1955. The number of matriculates was thus about 5300 per million literates; and the number of first graduates about 900 per million literates in India. In contrast, the number of first graduates in UK is 450 per million literates or proportionately only half that of India.

30. One gets a general impression that the number of students in the university stage is increasing quite fast and is roughly doubling every five years. The rate of increase in the number of science students is, however, much smaller at all stages and the rate of increase of science students is also appreciably smaller than arts students at the intermediate and degree levels and is roughly the same at the stage of the master's degree. The position is thus not satisfactory from the point of view of scientific and technical man-power.

31. The fact that the "proportion of graduates to literates" in India seems to be roughly double that of the UK makes it possible that education at the university level in India is expanding in quantity at the cost of quality. This, I believe, is also the general impression among my scientific colleagues. It raises some serious questions. It is true that when there is a large increase in the number of students, the average quality would usually decrease. It is doubtful, however, whether this has been the only reason. We are aware that a very large number of natural and social scientists were appointed during World War II, and are still being appointed in Government posts at salaries much higher than those that can be offered by universities and scientific institutions. In consequence there is a great shortage of competent and experienced teachers in the universities. The expenditure on laboratories and scientific equipment has not kept pace with the expansion in the number of students and is extremely small in comparison with needs. The syllabus and courses of instruction are in many cases out of date. All this requires to be changed.

Science Teaching

32. It is clear there cannot be any broad-based advance of science and scientific research in India without making university teaching more efficient and more oriented to present needs than at present. With this problem is linked the question of efficiency of science education in the top classes of secondary schools, to which I shall merely draw your attention but which I have no time to consider. It is necessary indeed to formulate a comprehensive scheme of science teaching, from the upper classes of secondary schools upto post-graduate teaching in the universities, and implement the same, in order to lay a sound foundation for science in India. The scientists have a very special responsibility in this matter. It is true, of course, that it would be necessary for Government to make important decisions of policy and to provide adequate funds. But, on the technical side, improvements in science teaching and research can be brought about only through vigorous and untiring efforts of scientists themselves. They have a triple task. They must take the initiative in formulating, at a concrete level and step by step, programme for the improvement of science teaching and research; and, secondly, persuade Government to take appropriate action in financial and policy matters. Thirdly, scientists should start doing whatever they can on their own; and then, as and to the extent Government support begins to be available, make the best use of such support for the promotion of science in India.

Scientific and technical man-power

33. It is generally agreed that India would require a rapidly increasing number of scientists, engineers, and technologists in future. It is necessary therefore to attract a sufficient number of young men of ability to take up science as a career. The only way is to offer adequate economic and social incentives. Consider the position in the advanced countries of the world. USSR has established a definite lead over all other countries in scientific and technical personnel. Professors and research scientists, as a group, get the highest salary and have the highest status in USSR. The Academicians (members of the Academy of Sciences) have a life pension of 5000 roubles (half of which would accrue to a surviving widow) in addition to the salary of the post. It is being increasingly realized all over the world that it is this high social appreciation of scientists which is the real secret of the rapid advancement of science in USSR. In USA, UK and other countries of Europe, professors and scientists have a high social status but the position is not as satisfactory as in USSR, which has become a matter of serious concern to these countries.

34. The position of science professors and research scientists in India is most unsatisfactory compared to advanced countries. Scientists working in universities and in scientific institutions receive much less salary than scientists working under Government. Scientists in Government employment also are in an unsatisfactory position. As a group (with a few exceptions) scientists receive less pay and have lower status than administrative officers. Also, scientists have very little say in the making of decisions in Government. An urgent requirement for the advancement of science in India is to remove such disparities.

SCIENTIFIC SOCIETIES AND PLANNING

35. I should now like to draw attention to another problem which is likely to become increasingly serious in future. In the Western countries scientific research and science teaching had developed at first in the universities, and later, in connexion with research in industry independently of Government control. From the 17th century many scientific societies and institutions were established in the countries of the West for the promotion of science and scientific research. More recently Government began to sanction large grants for scientific research. In most countries the largest share of research funds now comes from Government.[1] This has raised important questions of policy in regard to Government control over scientific research.

36. It is interesting to observe that in USSR an Academy of Sciences was established by the Communist Government at a very early date as an autonomous institution independent of the executive control of Government. The Academy has been given great responsibilities and supplied with adequate funds for a wide range of functional activities in scientific research and the promotion of science. It maintains over one hundred scientific institutes spread over the country, and in 1957 had a total staff of 50,000, of whom 12,000 were professional scientists and 4,000 professors. The point to be emphasized is that in all advanced countries of the world, in the West as well as in USSR, there are well established scientific societies with a great tradition of scientific thinking and criticism which are outside Government and which can be relied upon to supply Government with independent advice in scientific matters.

37. It seems to me that there is need in India of functional scientific institutions which would be outside Government but which would take up, by agreement with Government, such scientific work or research or promotion of science or the coordination of scientific activities as can be done conveniently and efficiently by a non-official agency with the help of a whole-time staff of competent and high level scientists. Scientists engaged on such "decentralized" research or scientific activities, on a whole-time salary basis, would be independent of Government. In this way it would be possible to build up an organization or system of scientific thinking and criticisms which would be able to render great service to Government by offering unbiased and independent views in scientific matters. There is great need of such advice in connexion with planning and in the appraisal of the progress of plan projects.

38. To speak quite frankly, some of these ideas have come to mind, because I am not satisfied with the present activities of the National Institute of Sciences. I feel scientists must shoulder greater responsibilities in connexion with national planning. Coming together for occasional meetings is not enough. I should, therefore, like to see the National Institute of Sciences undertake assignments relating to planning for national

---

[1] In USSR, China, and other socialized economies all expenses for research are provided by Government. In other countries also, scientific research is being increasingly financed by Government. Even in the USA, out of the total expenditure of \$ 5.4 billion dollars (= Rs. 2550 crore) in 1953, the share of Government was highest and 52 percent. Industry came next with a contribution of 44 percent. The share of universities was only 3 percent and that of institutes only one percent. Of the total expenditure, the biggest portion (72 percent) was spent by industrial concerns, 18 percent by Government agencies, and only 9 percent by universities. It is interesting to note that of the money spent on research by industry, 60 percent came from industry itself and 40 percent from Government. (*Scientific American*, November 1957, p. 47).

development which would be done in a systematic and painstaking way by a whole-time or part-time paid staff. I have a feeling that it is only through work at a concrete level that we shall be able to develop seriousness of purpose and a sense of responsibility in scientific matters of importance for national development.

Strategy of planning in India

39. The task of planning must be to improve continually the level of living, especially, of the 95 percent of the population who are poor, and to create enough new employments to get rid of the fear of unemployment. It has been recognized for a long time that this can be done only through rapid industrial development. As India has plenty of iron ore, coal, bauxite, etc., the key to industrialization lies in establishing the manufacture of heavy machinery, heavy electrical equipment (turbines, generators, switch and transmission gear), and machine tools. Once this is done, everything else can be gradually manufactured in India mostly out of domestic resources. As more and more machinery becomes available, it would be possible to create more and more employment and continually to construct more machinery and expand the production of coal and electricity; steel and aluminium ; fertilizers, cement and heavy chemicals; equipment for transport and communications; scientific instruments, etc.; and, in this way, produce more and more consumer goods and improve the level of living.

40. It has to be accepted, in principle, that until sufficient machinery and tools become available, it is essential to utilize idle hands for the production of cloth, household utensils and other consumer goods. This is for two reasons. Our greatest shortage is of machinery, while we have millions of unemployed persons sitting idle. Whatever machinery we can afford to import or manufacture within the country should be used to build up the basic industries which would produce more machinery, more steel, cement, fertilizers, electricity, etc. That is, we must not waste our meagre resources of foreign exchange to import non-essential consumer goods, or machinery for the manufacture of consumer goods which can be produced within the country by putting idle hands to work. Expansion of small scale and village industries would create employment all over the country, increase the supply of cloth and other consumer goods made by hand labour, and improve the level of living without hampering the expansion of basic industries.

41. A third point requires to be recognized. Agriculture and industry are closely interlocked in India. Industrial development is not possible without an increasing supply of cheap food and raw materials. On the other hand, in India, owing to scarcity of land, basic improvements in agriculture can be brought about only through the use of more fertilizers, more irrigation, and better methods of cultivation, all of which would have to depend on large scale industrial development. However, in the short run, a good deal can and must be achieved by intensive cultivation of land by hand and by improving conditions of living in rural areas through community projects, land reform, consolidation of holdings, village cooperatives, etc.

42. This brings us to the fourth point, namely, that planning in India must proceed at two different and clearly distinguishable levels or spheres. One is concerned with questions of large scale industrial development which can be, and are, decided in every country by a small group of persons at the top. For example, the decision in 1949

to instal a million-ton steel plant, its reversal in 1950, and fresh decisions in 1953 and 1954 to expand steel production by 3 or 4 million tons were all taken by the Government of India without any consultations with the general public. It is proper for Government to make such decisions. It is also possible to implement the decisions without any difficulty as it would involve only a small number of persons.

43. The position is entirely different in agriculture or in small scale and hand production which involve millions of households scattered over a very big country. Consider consolidation of holdings or village cooperatives. It is necessary, of course for Government to make policy decisions; but this by itself would not be sufficient because the implementation would depend on securing the willing cooperation or at least the concurrence of millions of persons in the villages. In the beginning, the organization of agriculture and small scale production, which may be called the "diffuse" sector, would be far more difficult than establishing large scale industries. The approach must be through institutional changes, spread of education, improvement of communication, etc., and, above everything else, by carrying conviction to the masses that Government is doing its best for national development. Success in establishing the basic industries would give Government increasing control over the national economy, and thus place Government in a strong position to make steady improvements in the diffuse sector and to secure the active support of the masses for national development.

44. Basic industries are thus of crucial importance in two different ways. Firstly, these alone can give us economic independence. Secondly, these alone can place Government in a position to make steady advance in the diffuse sector. The strategy of planning, therefore, requires the highest priority being given to the basic industries, namely, heavy machine building, heavy electrical equipment, machine tools, fertilizers, steel, cement, etc.

The Question of Foreign Exchange

45. It is true that, in the beginning, it is necessary for us to have both technical and financial help from outside India. Professor P. M. S. Blackett, in his Presidential Address to the British Association for the Advancement of Science in September 1957 has pointed out the importance of outside help to an underdeveloped country in getting it off on its way to economic development. He has also given an inspiring lead by stressing the responsibility of the industrially advanced country in this matter from the world point of view. I fully appreciate the need of external aid.

46. I am also aware that grave difficulties have been created by the shortage of foreign exchange. A new approach was adopted in the Second Five Year Plan, and the size of the Second Plan was intentionally made much bigger than that of the First. It is not surprising that mistakes have been made at the stage of transition from a small to a much bigger scale of planned economy. Techniques and tools of planning are not yet adequate and would have to be improved; and these questions are receiving the attention of Government and the Planning Commission. It would be appropriate, however, to examine the question of foreign exchange from the point of view of long-term planning.

47. I shall give a numerical example. Consider the question of foodgrains about which India is in a precarious position. From 1948 to 1953 we imported 19.3 million

tons of cereals at a cost of about Rs. 867 crore in 6 years, that is, at the average rate of about 3.2 million tons at a cost of Rs. 145 crore per year.[1] The average price paid was about Rs. 450 per ton.

48. The population of India is growing at the rate of, possibly, 5 million persons per year.[2] It takes roughly one ton of cereals to feed 7 persons. At this rate, an additional supply of 700,000 tons of foodgrains *every year* would be required to keep pace with the growth of population. If we decide to purchase this quantity from abroad, we would require *every year* an additional Rs. 30 crore of foreign exchange which would amount to Rs. 150 crore in the course of one five-year plan (and this, also, only at the present rate of growth of population).

49. Consider the alternative method of using imported fertilizers to increase the production of cereals. One ton of ammonium sulphate, at a cost of about Rs. 250 per ton, should normally give an additional yield of 2 or 2.2 tons of foodgrains. To give 700,000 tons more of foodgrains, the quantity of ammonium sulphate required would be 350,000 tons, of which the price would be Rs. 8.75 crore or say Rs. 9 crore. This would be the additional amount required *every year;* in the course of 5 years the total amount required would be Rs. 135 crore.

50. A still better way would be to instal *every year* a new fertilizer factory to produce 350,000 tons of ammonium sulphate (roughly, like the factory at Sindri). The fixed investment would be about Rs. 25 crore, of which the foreign exchange component would be only Rs. 10 crore. After the period of construction is over, each such factory would supply enough fertilizer to take care of the increase of population of one year. The total expenditure in five years would be Rs. 125 crore with, however, a total foreign exchange component of only Rs. 50 crore.[3]

51. It is possible to go one step further, and establish immediately a heavy machine building factory in India, and, when this plant is constructed, to manufacture in India the machinery required, to set up a fertilizer factory. Of course, it is neither possible nor necessary to set up a machine building plant exclusively for the manufacture of big machinery for fertilizer factories; the work can be done in plants which would manufacture machinery for other industries as well. It is, however, possible to make rough calculations of cost, because in the case of heavy machine building plants the value of output is roughly equal to the value of the fixed investment. To produce machinery worth Rs. 10 crore (the value of imports required to set up a Sindri-type plant) would thus call for an investment of about Rs. 10 or, say, 15 crore with a foreign exchange component of

---

[1] Imports fell below one million ton in 1954 and 1955 but we have again started importing foodgrains heavily from 1956.

[2] There is some evidence to indicate that the rate of growth may be appreciably greater and as high as 6 or even 7 million per year. Although registration of births and deaths is known to be extremely defective, the fall in the recorded death-rate is probably real. Population estimates based on the National Sample Survey show high rates of growth in some of the rural areas.

[3] Besides this, there will, of course, be current costs of intermediate products used up and wages paid. But when the intermediate products are available within the country and the output of wage goods industries is sufficient to meet the demand, such costs could be met entirely out of domestic resources. This would, no doubt, call for an expansion of the economy and would require organizational efforts; but similar and even greater organizational efforts would be needed to produce the goods to be exported to meet the larger requirements of foreign exchange in the other cases.

possibly Rs. 8 or 10 crore. This would be the (apportioned) cost in the first instance. The actual cost of setting up a fertilizer factory would be Rs. 25 crore per year or Rs. 125 crore in 5 years ; but this would not involve any expenditure of foreign exchange.

52. To sum up, in order to feed the fresh additions to the population at the rate of 5 million persons per year, it would be necessary to provide an additional quantity of 700,000 tons of foodgrains *every year* which would require Rs. 450 crore of foreign exchange over a period of five years. The cost can be reduced to Rs. 135 crore of foreign exchange in a 5-year period, if an additional quantity of 350,000 tons of ammonium sulphate is ordered from abroad *every year*, at least two years in advance of the crop season. The cost can be further reduced to Rs. 125 crore (out of which the foreign exchange component would be Rs. 50 crore) over a five-year period, if a new fertilizer factory of 350,000 ton capacity is started *every year*; this would call for decision 4 or 5 years ahead of the crop season concerned. The apportioned cost of a heavy machine building factory which would manufacture machinery in India to instal *every year* a new fertilizer factory of 350,000 ton capacity[1] would be, however, so small as Rs. 12 or 15 crore with a foreign exchange component of perhaps Rs. 8 or 10 crore. Such a decision would have to be made only once but 8 or 10 years in advance of the season in which the fertilizer would be used.

53. The above account is, no doubt, over-simplified ; but it brings out clearly one important point. In India it is of the greatest importance to take a long view of planning. Additional requirements of foodgrains to cope with the growth of population can be met with an (apportioned) investment of only Rs. 15 or 20 crore and of foreign exchange of Rs. 8 or 10 crore, provided a decision can be made 8 or 10 years in advance of the year in which the food would be consumed. If the decision is made 4 or 5 years in advance, the cost would increase to Rs. 125 crore with a foreign exchange component of Rs. 50 crore. If the decision is made a couple of years ahead, the cost would increase to Rs. 135 crore in foreign exchange to import fertilizers; and if the decision to import foodgrains is made only from year to year the cost would go upto Rs. 450 crore in foreign exchange. *Wise utilization of Rs. 10 crore of foreign exchange once for all can lead, in time, to a saving of Rs. 450 crore of foreign exchange in imports of foodgrains over a period of 5 years.*

54. Similar considerations hold good in other crucial sectors. An investment of Rs. 150 crore, including Rs. 80 crore of imported machinery, would be required to instal a million-ton steel plant with a product value of Rs. 40 or 45 crore. A heavy machine building factory with an investment of Rs. 80 crore, with foreign imports worth about Rs. 50 crore, would produce every year machinery worth roughly Rs. 80 crore or the equivalent of imported machinery needed to set up a million-ton steel plant. Once such a heavy machinery factory gets into production, it would be possible to start a new million-ton steel plant *every year* out of our own resources. An investment of Rs. 30 crore in a plant to manufacture mining machinery would produce goods worth about Rs. 30 crore per year. An investment of Rs. 100 crore in heavy electrical equipment (generators, switchgear, heavy motors, rectifiers, transformers, etc.) would give products of the value of about Rs. 70 crore per year; and so on.

---

[1] There is great scope for the use of fertilizers in India ; some relevant information is given in Statistical Notes on the Use of Fertilizers.

Need of Long-term Planning

55. From the point of view of industrial development over a period of 15 or 20 years there cannot be any doubt that the correct policy would be to establish the heavy machinery industries. Consider, purely for purposes of illustration, the following programme of investments with a rough value of the product given within brackets in each case :—heavy machinery Rs. 150 crore (150 crore), heavy electrical equipment Rs. 100 crore (70 crore), turbines Rs. 20 crore (12 crore), mining machinery and drilling equipment Rs. 30 crore (30 crore), machinery for chemical engineering Rs. 25 crore (25 crore), machine tools[1] Rs. 30 crore (20 crore), aluminium Rs. 25 crore (5 crore), and, say, another Rs. 70 crore for other basic industries giving a total investment of about Rs. 450 crore with foreign exchange requirements of the order of Rs. 300 crore over a period of five years. I am aware that foreign exchange would also be needed to expand transport and other facilities but provision has been already made for this and for steel and other investments in the Second Plan.

56. The point I am emphasizing is that Rs. 300 crore of foreign exchange invested in suitable basic industries over a period of 5 years would establish a sound foundation for future industrial development and would enable us to manufacture essential capital goods within the country. (I am not suggesting that we should reduce our imports. On the contrary, we should try continually to expand our foreign trade. But we must attain as soon as possible a position in which, if necessary, we shall be able to manage on our own in an emergency). We have made some real advance during the First Five Year Plan. It has been a very wise and farsighted decision to develop atomic energy, for which we have much natural resources. Dr. H. Bhabha has shown in a recent article[2] how great would be the benefit of atomic power electricity to Indian economy. Good progress is being made with steel. It is now necessary to make similar farsighted decisions to establish the basic heavy industries.

57. I should like to refer in this connexion to some recent trade statistics. Indian exports f.o.b. amounted to Rs. 597 crore in 1954-55; Rs. 641 crore in 1955-56; and Rs. 637 crore in 1956-57. Imports c.i.f. in these three years were respectively Rs. 684 crore, Rs. 757 crore, and Rs. 1077 crore. It would be noticed that the increase in imports in 1956-57 over 1955-56, that is, in one year was Rs. 326 crore; and in 1956-57 over 1955-54, that is, in two years was Rs. 393 crore, out of which the increase on private account was Rs. 272 crore. I may also mention that the Government outlay on the First Five Year Plan was nearly Rs. 2400 crore; and the outlay on the Second FYP is expected to be roughly double. It is in the context of the above figures that one must view the proposal to invest Rs. 450 crore and utilize Rs. 300 crore of foreign exchange to build up the basic industries. After the factories come into operation, there would be a saving of Rs. 300 or Rs. 350 crore of foreign exchange in the import of capital goods. This is the crucial issue of planning in India.

---

[1] To give one example, it is worth mentioning that machine tools formed roughly 3 percent of industrial equipment (in value) over a long period in the USA. On this basis Rs. 20 crore worth of machine tools would suffice for the manufacture in India of about Rs. 600 crore worth of industrial equipment which, in its turn, would enable industrial investments being made to the extent of *at least* Rs. 1000 crore per year basically out of our own resources. (*The Machine Tool Industry*, New York, Clark, Dodge & Co., May 1948).

[2] *Science and Culture*, October and November, 1957.

58. Finance and balance of payments are important aspects of planning; and there is no conflict between financial planning and physical planning. But it must be kept in mind that, in an underdeveloped country, the physical view is of basic importance. In one single year, 1956-57, our imports suddenly exceeded exports by Rs. 440 crore. If by some happy chance, Rs. 300 crore out of this amount had been utilized to purchase capital goods required to establish the basic industries, then in five year's time we would have won our economic independence and would have overcome difficulties of foreign exchange permanently. Although the accounting position would have been the same, the real position would have been entirely different and ever so much better.

Planning in China

59. It is instructive to consider what happened in China. Mr. Li Hsien-nien, Vice-Premier and Minister of Finance of China, in his speech to the First National People's Congress in Peking on 29 June 1957 (supplement to *Hsinhua News Agency Release, Peking*) stated that China has received foreign loans amounting to 2,174 million yuan (= Rs. 435 crore) before 1953, and 3120 yuan (= Rs. 624 crore) for China's First Five Year Plan from 1953 to 1957. The total amount of the loan was thus Rs. 1059 crore. A good part of this was used for rehabilitation after the present Government came into power (as stated by the Finance Minister); and a part, no doubt, was used for military expenses during the war in Korea. However, with the whole of the second loan of Rs. 624 crore and a part of the first loan China rapidly developed the basic industries (heavy machinery, heavy electricals, machine tools, steel, fertilizers, trucks, etc.). The Finance Minister of China appreciated the help received from USSR, but stated "*we may say that we are now in a better position to rely on our own accumulation* [i.e., investment] *to carry on national construction.*" (P. 5, col. 1).

60. It is important to note that foreign credit had tapered off rapidly. In the same speech the Finance Minister stated that foreign loans constituted 117 million yuan or Rs. 23.4 crore in 1956, and 23.3 million yuan or only Rs. 4.7 crore in 1957. China has a foreign trade which is roughly of the same size as India on a per capita basis. Although foreign loans would be of help and would make the task lighter, it seems that China will not have to worry about foreign exchange for essential requirements in future, and would be able to manage, if necessary, on the trading surplus. It may be mentioned that India received about Rs. 200 crore of foreign assistance, roughly half in grants and half in loans, during the First Five Year Plan; and has fairly firm authorization of receiving at least Rs. 600 crore of foreign assistance during the Second Five Year Plan. India also had foreign assets (held by the Reserve Bank) to the extent of Rs. 723 crore in 1951-52 and Rs. 746 crore in 1955-56 at the end of the First Five Year Plan which China had lacked. The foreign exchange position in India was thus much more favourable.

61. In Chinese planning the highest priority was given to develop those basic industries which would enable China to manufacture, within the country, essential capital goods required for rapid industrial development in future. Remembering that the Chinese economy is bigger than ours, it is of interest to note that China has "got off" on the way to industrialization out of its own resources with comparatively less foreign exchange

than India. In contrast, India may continue to have foreign exchange difficulties for some considerable time to come in the absence of wise planning.[1]

62. The Chinese example merely corroborates what an objective scientific analysis had brought out, namely, that India could have built up, and can still build up, with the help of only a few hundred crores of foreign exchange those basic industries which would give her economic independence.

Selfconfidence and Economic Independence

63. It is our responsibility, as scientists, to work out the logic and techniques of planning suited to India. It is our responsibility, as scientists, to educate and convince our countrymen that India is a great country with vast resources, natural and human. We have the responsibility to use these resources to improve the level of living of our countrymen and to promote science and culture. We should welcome foreign help and cooperation, especially, in science and technology and in cultural matters. Also, we must realize that it is not necessary continually to depend upon help from abroad. We must overcome the defeatist attitude which would make us continually ask for foreign loans. We must realize that political independence is not enongh; we must win economic independence. We have the ability and the responsibility to make rapid economic advance with our own resources, and with hard labour, scientific thinking, wise decisions, and seriousness of purpose. We must have faith in ourselves.

---

[1] The Indian Industrial Delegation which visited USA, Canada, UK, West Germany and other countries in Europe in September–November 1957, referring to foreign investments in Canada in its Report to the Federation of Indian Chambers of Commerce and Industry, stated :

> "Eight hundred dollars per capita foreign investment means nearly Rs. 4,000 in Indian currency and at that rate and for such high development India should require Rs. 1,60,000 crore (Rupees one lakh sixty thousand crore) of investment in India, of which at least one-third, viz., nearly Rs. 50,000 crore would have to be in foreign currency! This can only be a dream. The conclusion to draw is that India cannot be developed without foreign capital which we shall continue to need for at least the next 25 years and in substantially large amounts" (p. 27).

This attitude of mind is in sharp contrast to the outlook in China where, with a fifty percent bigger population than India, the national economy got off on the way to rapid economic development with the help of foreign loans of only Rs. 1059 crore spread over 7 or 8 years.

Statistical Notes

Notes on Table 1 : In Table 1 figures have been given in metric tons. The Chinese official estimate of population in 1955 has been given instead of the UN estimate. Electricity production in China and India excludes generation in enterprises for their own use. Figures of energy (all forms) presented in col. (7) refer to the gross inland consumption of commercial fuels and water power, expressed in terms of coal equivalent. Production data include chiefly coal and lignite, gasoline, kerosene and fuel oils, natural gas, and hydro-electric power; electricity produced by using the above fuels has been excluded. No account has been taken of fuel wood and other vegetable fuels and peat for want of adequate data. These possibly provide a major share of the energy supply in all but a few countries. In India, for example, a United Nations source (*World Energy Supplies in Selected Years, 1929-50*) gives a figure for 1949 of about 84 million tons in coal equivalent of energy from such sources (mostly dung and fuel wood) compared to a figure of 35 million tons in coal equivalent for all commercial sources of energy quoted in Table 1. Thus we get a ratio between 2 and 3 to 1 between non-commercial and commercial sources in India. The corresponding figure for USA is only 1 is to 20. For India, the percentage distribution of all energy may be given as : commercial energy (19%); non-commercial sources such as fuel wood, dung, etc. (44%); animal effort (33%); and human effort (4%) with total (= 100%). The UN publication used above did not include the last two sources.

Notes on Table 2. In Table 2 the national income figure for USSR is calculated from an estimate given in the *Economic Bulletin for Europe*, May, 1957 on the assumption that one-third of the gross savings comprise depreciation. Other figures have been taken from sources referred to in the footnote to the table and have been rounded off. Also, all the figures have been expressed in national currencies. The conversion factors between the various currencies used are 1 $ = 4.75 Rs.; 1 £ = 13.3 Rs.; 1 rouble = 1.2 Rs. and 1 Yuan = 2 Rs., at current official exchange rates. The exchange rates and the figures in col. (3) would give some indication of the wide difference in the level of living between advanced and underdeveloped countries; any finer comparison between per capita incomes is fraught with grave dangers of error because of structural dissimilarities in prices.

The per capita figures in Table 2, col. (4) are obtained by using the output and population data in Table 1, and figures in col. (5) are directly derived from *UN Statistical Year Book* 1956 in terms of coal equivalents (p. 308—10); the conversion factor used being 1 ton of coal = 1667 kwh of energy.

Greater difficulties are associated with the derivation of the expenditure on research given in col. (6) because of inadequacy of information. The figure of $ 5.4 billion dollars used for USA relates to 1953 and is on the basis of a report on *Funds for Basic Research in the United States* 1953 published by the National Science Foundation, USA; and also briefly reported in the *Business Week*, 23 November 1957. It was noted that only about 8 percent of the amount is devoted to fundamental research. The percentage for USA given in col. (7) is computed on the basis of national income for 1953.

The expenditure on scientific research in USSR is taken as 18.2 billion roubles on the basis of a report in the *Times of India* of a message dated Moscow, 20 December 1957, which was later confirmed in other communications.

The expenditure on scientific research in the UK is taken to be £325 million on the basis of an article, "Highly trained technical man-power in Western Europe," *Economic Bulletin for Europe*, May 1957.

The figure used for scientific research in China is 327 million yuan (= Rs. 65.4 crore approximately) on the basis of the Finance Minister's speech on the budget for 1957 (July 1957). It should be noted that this expenditure is on scientific research proper and excludes a considerably larger figure (637 million yuan = Rs. 127.4 crore) for higher education. However, we have to take cognizance of the fact that in China and USSR the definition of 'science' is wider and includes some of the social sciences (economics, linguistics, history, etc., which are not reckoned as science in Western countries) but excludes art, literature, etc. This would give a higher total figure for research in socialist countries. On the other hand, much of developmental research in engineering and technology is probably included under industries in socialist countries.

The figure for expenditure on scientific research in India has been arrived at on the basis of Government budget accounts and financial statements. Attempts have been made to exclude expenditure on higher education and certain promotional outlays ; but expenditure on research in social sciences, such as economics and statistics has been included. No account, however, has been taken of research met exclusively by bodies other than Government; but its magnitude is small in India. The figure of Rs. 15 crore used is, at best, a dimensional estimate.

Exact comparisons are not possible because the definition of "scientific research" for budget and accounting purposes varies from one country to another. It is, in any case, difficult to demarcate research in applied science, in a precise and comparable way, from developmental work in engineering and technology. The figures given, however, give a broad idea of the position in the different countries.

Notes on Table 3. *USA :* Data on available number of engineers and doctors as well as the figure of annual out-turn of doctors are taken from the *Statistical Abstract of USA*, 1957. The figure for annual out-turn of engineers is quoted from "Trained Man-power, USA *vs.* USSR," *Monthly Review*, New York, April 1957.

USSR : Regarding engineers, both the figures are based on Soviet official statistics (in Russian). The figure of number of doctors available is taken from the English language publication, *National Economy of USSR : Statistical Returns*, while the out-turn figure is from the *USSR Reference Book*, November 1957.

UK : Both the figures for engineers are taken from "Highly trained technical man-power in Western Europe," *Economic Bulletin for Europe*, May 1957. The Annual Report of the British Medical Association, 1955-56 (Supplement to the *British Medical Journal*, April 7, 1956) gives the total membership of the Association as 68,306 in March 1956 and states that this number represents 80% of the total number in the profession ; the available number of doctors is estimated on this basis. The figure of annual output is obtained from G.D.H. Cole : *Post-War Condition in Britain*, 1956.

China : All the figures are based on the statistical returns for the People's Republic of China prepared by the State Statistical Bureau (in Chinese). However, a base figure of the available number of engineers in 1952 has been taken from Solomon Adler : *Chinese Economy*.

India : The out-turn figures are based on the *Educational Statistics of India*. The available number of engineers is taken from *Man-power Studies No. 5 : Engineers in India : Number and Distribution 1955*, (Planning Commission, October 1957). The number of doctors is based on information available in the Planning Commission.

Notes on the Use of Fertilizers : In a review by the Indian Council of Agricultural Research on "The Use of Fertilizers on Foodgrains"[1] the response, standardized to 20 lbs. N (= 100 lbs. of sulphate of ammonia) per acre, was given as 3.4 maunds (or 274 lbs.) of rice per acre, 2.8 maunds (or 230 lbs.) of wheat per acre, and between 1.7 and 3.1 maunds (140 lbs. and 255 lbs.) of other cereals per acre . For paddy the water supply was presumably adequate. Experiments were reported for wheat from both irrigated and unirrigated crops; the response on unirrigated land was 1.7 maunds (140 lbs.) per acre compared to 2.8 maunds (230 lbs.) per acre on irrigated land. The authors pointed out that

> "when the total supply of fertilizer is severely limited, the greatest natural benefit will accrue if small dressings are used on a large area of crop rather than heavy dressings on a small part of the area. ....The first step....is to ensure that available supplies are distributed over a large area of potentially responsive land; the second step is to increase supplies so that all areas likely to benefit considerably can have a small dressing; only thirdly does the increase of dressings to optimal levels become a desirable national policy" (p. 14).

The authors concluded that nitrogen response of 2.5 maunds (= 206 lbs.) of grains per acre from 20 lbs. N per acre (= 100 lbs.sulphate of ammonia) was "a reasonably representative figure for all cereals" (p. 15).

---

[1] By F. Yates, D. Finney and V. G. Panse : I.C.A.R. Research Series No. 1, 1953.

From information supplied by I.C.A.R. for a later (T.C.M.) series of field trials on rice and irrigated wheat conducted during the three years 1953-54 to 1955-56 at community project centres distributed over the whole country, it appears that the nitrogen response for a first dose of 20 lbs. of nitrogen per acre was about 3 maunds (247 lbs.) of both rice and wheat per acre; and for a dose of 40 lbs. of nitrogen per acre the response was 4.5 maunds (370 lbs.) of rice per acre and 4.7 maunds (380 lbs.) of wheat per acre.

There is a gradual falling off in the response with higher doses of nitrogen; but from the results given above it is clear that upto about 20 lbs. of nitrogen (100 lbs. of sulphate of ammonia) a response of 200 lbs. or twice by weight of foodgrains can be expected on an average with confidence.

The net area sown in India was 313 million acres and the gross area sown including double cropped land was over 351 million acres in 1953–54.[1] The gross area irrigated was about 60 million acres. The total area under rice is about 78 million acres, out of which about 26 million acres are irrigated leaving a balance of about 52 million acre of unirrigated rice land, much of which would normally have adequate rainfall. We thus have about 34 million acres of irrigated land under crops other than rice, 26 million acres of irrigated rice land, and 52 million acres of unirrigated land under rice giving a total of 112 million acres of land, for which the water supply is likely to be fairly adequate for the use of fertilizers. At a rate of application of 20 lbs. of nitrogen per acre there should not be any difficulty in using upto one million tons of nitrogen or 5 million tons of sulphate of ammonia for this type of land alone. This is a safe estimate of minimum requirements. For the remaining 200 million acres of land if we accept a comparatively small doze of 10 lbs. of nitrogen per acre on an average, we could use nearly the same quantity again at the second stage. This would raise the requirement to about 10 million tons of sulphate of ammonia per year. For 112 million acres of land with good supply of water it would be possible to increase the doze upto 40 lbs. of nitrogen per acre which would require nearly another 5 million tons of sulphate of ammonia. These very rough calculations indicate that India may be able to use without any difficulty upto 15 or 16 million tons of sulphate of ammonia per year, that is, can afford to instal 40 or 45 factories of the Sindri capacity of about 350,000 tons per year.

We may also consider the question from the point of view of rainfall. There are about 50 million acres of land under cultivation which get more than 50 inches of rainfall per year; and about 87 million acre under cultivation with an annual rainfall of roughly between 40 and 50 inches. Out of these 137 million acres about 36 million acres are irrigated leaving a balance of about 100 million acres of unirrigated land with annual rainfall of about 40 inches or more. The balance of irrigated land is about another 30 million acres giving a total of about 130 million of acres under cultivation which is likely to have enough water for the profitable use of good doses of fertilizers. On the other hand, only about 50 million acres of land under cultivation in India receive less than 30 inches of rainfall per year; and of this about 12 million acres are irrigated. This would leave something like 38 million acres of land which may be really unsuitable for the use of fertilizers.

I may further mention that the rate of application of nitrogen in many countries is very high.[2] For example, the natural average doze of nitrogen in lbs. per acre over all cultivated area in 1955-56 was as follows: Belgium 73, East Germany 42, West Germany 48, Netherlands 156, Norway 38, U.K. 37, Japan 98 and Egypt 38 compared to 0.86 lbs. per acre with a total consumption of 145,000 tons of nitrogen or 725,000 tons of sulphate of ammonia in India. These figures also indicate that there is great scope for the use of fertilizers in India. The proposal made in my Address of installing only one fertilizer factory of the same capacity as Sindri per year is thus quite modest. It is worth mentioning in the present connexion that there is some evidence to suggest that the actual rate of growth of population may be appreciably greater than 5 million per year and may be 6 or 7 or even more per year. It would be wise to instal two such factories per year and this can be done without any fear of the saturation limit of utilization being reached in 15 or 20 years.

---

[1] *Agricultural Statistics of Reorganised States,* Government of India, 1956–57.

[2] Year Book of Food & Agricultural Statistics: Production, Vol. 10, Part I (FAO), 1956.

Statistical Appendix

Household consumer expenditure

1. Since 1950 the National Sample Survey (NSS) has been collecting a wide variety of social and economic statistics from all over the country. A great deal of information has thus become available on household expenditure on items of consumer goods such as foodgrains, sugar, milk and milk-products, or all food items taken together, clothing, medical expenses, etc., as well as the total expenditure on all items for a period of 30 days. Data collected in the NSS : 10th Round (December 1955—May 1956) for 2754 sample households have been used to show some interesting facts in Tables A.1.0, A.1.1, A.1.2 and A.1.3. For each household it is possible to calculate the "expenditure per person" by dividing the total expenditure on all consumer goods by the number of persons in the household. In this way each person in the sample would be labelled by the "expenditure per person" of the household to which he belongs. It is then possible to classify all the individuals included in the sample into a number of groups or classes of expenditure level. For example, in Table A.1.0, the individual persons were classified into groups having 'expenditure per person up to Rs. 8', 'between Rs. 8 and Rs. 11', 'between Rs. 11 and Rs. 15', etc., and finally 'Rs. 55 and above'. These class-ranges are shown in the top line of Table A.1.0 at the heads of cols. (1) to (13). Information on different items is given in different lines separately for each class-range in the corresponding cols. (1) to (13).

2. In the first portion of the table the average expenditure per person in rupees per 30 days is shown for 4 'food' items in lines 1 to 4; the total expenditure on food items in line 5; on clothing, medical expenses and 'others' in lines 6–8; non-food total in line 9; and finally the expenditure on all items of consumption in line 10.

3. The expenditure on each item can be expressed as a percentage of the total expenditure given in line 20. These percentages are shown in lines 11 to 20 and show many interesting features. The percentage expenditure on foodgrains (line 11) is over 55 percent in the poorest group in col. (1) and decreases to about 17 percent in the highest group in col. (12). The total expenditure on food (line 15) falls from 79 percent at the bottom in col. (11) to 44.8 percent at the top in col. (12) with an overall average of 65.4 percent for all families. This shows that for India as a whole, food items make up about two-thirds of the household expenditure; in the poorest section it can be as high as four-fifths. Many other comparisons can be made on the basis of the results given in this table.

4. The average size of households, that is, the number of persons in the household is given in line 23; the estimated number of households (in million) in line 24; the estimated number of persons (in million) in line 25. It appears that, according to the NSS, there were nearly 80 million households and 385 million persons in India at the time of the survey. The expenditure in Rs. crore per 30 days is also given in lines 26 to 28. The household budget of India came to about Rs. 722 crore per 30 days, out of which about Rs. 472 crore were spent on food and Rs. 250 crores on non-food items.

TABLE A.1.0. CONSUMER EXPENDITURE IN RUPEES PER PERSON FOR A PERIOD OF 30 DAYS WITH PERCENTAGE OF TOTAL EXPENDITURE : ALL INDIA (RURAL and URBAN) : 1955-56

items	expenditure classes in rupees per person per month												
	Rs. 0–8	8–11	11–13	13–15	15–18	18–21	21–24	24–28	28–34	34–43	43–55	55 & above	all classes
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
average expenditure in rupees per month													
1. foodgrains	3.49	4.69	5.89	6.35	6.93	7.58	8.26	8.52	8.71	9.33	10.08	13.98	6.61
2. sugar	0.09	0.20	0.25	0.36	0.42	0.52	0.65	0.77	1.05	1.13	1.46	2.41	0.51
3. milk & products	0.15	0.43	0.66	0.92	1.11	1.78	2.15	2.69	3.69	4.71	9.98	8.56	1.75
4. other food	1.22	1.85	2.34	2.73	3.28	3.54	4.31	4.81	5.67	6.01	7.43	12.34	3.39
5. food total	4.95	7.17	9.14	10.36	11.74	13.42	15.37	16.79	19.12	21.18	28.95	37.29	12.26
6. clothing	0.21	0.53	0.79	1.03	1.43	1.92	2.19	2.60	3.86	4.93	4.33	10.24	1.77
7. medical	0.01	0.09	0.11	0.19	0.27	0.37	0.46	0.51	0.84	1.21	1.73	3.90	0.43
8. others	1.09	1.62	1.94	2.38	3.05	3.83	4.49	5.89	6.86	10.40	12.23	31.86	4.28
9. non-food total	1.31	2.24	2.84	3.60	4.75	6.12	7.14	9.00	11.56	16.54	18.29	46.00	6.48
10. total (all items)	6.26	9.41	11.98	13.96	16.49	19.54	22.51	25.79	30.68	37.72	47.24	83.29	18.74
percentage of total expenditure													
11. foodgrains	55.75	49.84	49.17	45.49	42.30	38.79	36.69	33.04	28.39	24.72	21.34	16.78	35.27
12. sugar	1.44	2.13	2.09	2.58	2.55	2.66	2.89	2.99	3.42	3.00	3.09	2.89	2.72
13. milk & products	2.40	4.57	5.51	6.59	6.73	9.11	9.55	10.43	12.03	12.49	21.13	10.28	9.34
14. other food	19.48	19.66	19.52	19.55	19.88	18.12	19.15	18.64	18.48	15.93	15.72	14.82	18.09
15. food total	*79.07*	*76.20*	*76.29*	*74.21*	*71.19*	*68.68*	*68.28*	*65.10*	*62.32*	*56.15*	*61.28*	*44.77*	*65.42*
16. clothing	3.35	5.63	6.59	7.38	8.67	9.83	9.73	10.08	12.58	13.07	9.17	12.29	9.45
17. medical	0.16	0.96	0.92	1.36	1.64	1.89	2.04	1.98	2.74	3.21	3.66	4.68	2.29
18. others	17.42	17.21	16.20	17.05	18.50	19.60	19.95	22.84	22.36	27.57	25.89	38.26	22.84
19. non-food total	*20.93*	*23.80*	*23.71*	*25.79*	*28.81*	*31.32*	*31.72*	*34.90*	*37.68*	*43.85*	*38.72*	*55.23*	*34.58*
20. total	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
21. percentage of persons	14.01	17.31	11.74	9.14	11.55	8.21	7.04	5.53	5.17	4.85	2.59	2.86	100.00
22. cumulative percentage	14.01	31.32	43.06	52.20	63.75	71.96	79.00	84.53	89.70	94.55	97.14	100.00	
23. size of households (persons)	5.40	5.36	5.72	5.00	4.85	4.57	4.48	4.39	4.05	4.49	3.30	3.38	4.83
24. no. of households (million)	9.99	12.44	7.91	7.04	9.18	6.92	6.05	4.85	4.91	4.16	3.02	3.27	79.74
25. no. of persons (million)	53.96	66.67	45.21	35.21	44.50	31.62	27.12	21.32	19.91	18.68	9.99	11.03	385.22
total expenditure in Rs. crores per 30 days													
26. on 'food' items	26.69	47.83	41.33	36.48	52.23	42.43	41.68	35.79	38.08	39.58	28.92	41.14	472.18
27. on 'non-food' items	7.22	14.96	12.86	12.66	21.15	19.35	19.37	19.19	23.01	30.90	18.27	50.75	249.69
28. on 'all' items	33.91	62.79	54.19	49.14	73.38	61.78	61.05	54.98	61.09	70.48	47.19	91.89	721.87
29. no. of sample households	260	369	240	217	305	244	197	207	193	187	138	197	2754

no. of sample villages = 1544, no. of sample urban blocks = 1220

Based on the National Sample Survey (NSS) : 10th Round, December 1955—May 1956, All-India (rural and urban) excluding Jammu & Kashmir.

5. Line 21 gives the percentage of persons in each expenditure class. From this line it can be seen that 14.01 percent of the population spend less than Rs. 8 per person in 30 days. The next line 22 gives the accumulated percentages which is obtained by adding the figures in successive columns. This is very convenient, because it makes it possible to see directly the percentage of the population spending less than a specified amount. For example, from col. (2) it would be seen that 31.32 percent or almost a third of the population spend less than Rs. 11 per person per 30 days; from col. (4), that 52.2 percent or more than half the population spend less than Rs. 15;[1] from col. (9), that nearly 90 percent of the population spend less than Rs. 34.[2] This brings out the actual level of living of 90 percent of our countrymen.

6. The above information has been shown in a graphical form given in the chart in "Curve A on p 76: Expenditure on all items in rupees" (per person per 30 days) in a solid line in the upper left hand corner of the chart. The vertical scale on the left gives the cumulative percentage of the population and the horizontal scale at the top (in solid line) gives the scale of expenditure in rupees per person per 30 days. It is easy to read off from this graph the percentage of persons having an expenditure below any assigned value. From this it is also possible to obtain (by subtraction from 100) the percentage of persons having an expenditure greater than any assigned value. It is, of course, also possible to obtain for any assigned percentage the limiting value of the expenditure. These values are given in Table 3 in the main text.

PERCENTAGE EXPENDITURE ON CONSUMPTION

7. The information given in Table A.1.0 is shown in a different form in Table A.1.1. In each line the average expenditure for all classes, given in col. (13) of Table A.1.0, is adopted as the basis for comparison and is put=100. Each figure in the same line is then expressed in the form of a percentage of the figure in col. (13). These are given in the form of index-numbers in Table A.1.1 and show how the expenditure on each item and the value of other quantities increase with an increase in the level of expenditure.

8. For example, in line 1 for foodgrains, it is seen that in col.(1) the expenditure on food is 53 percent of the average expenditure for all classes taken together. In contrast, in col. (12), the highest group, the index-number is 211 showing that the relative expenditure on foodgrains is double that of the general average. Also, comparing the index-number 211 with the index-number 53, it is seen that the expenditure is four times greater in the top group compared to the bottom group. For sugar (line 2), the bottom and top index-numbers are 18 and 473 respectively showing an increase of 26 times. Similar comparisons can be made for other items; but, because of errors of sampling, the comparisons should not become too refined. A glance at col. (12) would also show that expenditure in the top group is relatively much greater on non-food items (710) compared to food items (394), and very high (907) in the case of medical expenses. In fact, medical care would seem to be the highest luxury in India which only the richer people can afford.

---

[1] Approximately $3 (three US dollars) and British 22s. 6d. or US 10 cents and 9 pence per head per day.

[2] Approximately $7 (seven US dollars) and British 51 shillings or less than US 25 cents and British 1s. 6d. per head per day.

TABLE A.1.1. INDEX NUMBER OF CONSUMER EXPENDITURE PER PERSON FOR A PERIOD OF 30 DAYS BY ITEMS OF CONSUMPTION : ALL INDIA (RURAL AND URBAN) : 1955-56

items	expenditure classes in rupees per person per month												
	0–8	8–11	11–13	13–15	15–18	18–21	21–24	24–28	28–34	34–43	43–55	55 & above	all classes
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
Index number of average expenditure per month													
1. foodgrains	53	71	89	96	105	115	125	129	132	141	152	211	100
2. sugar	18	39	49	71	82	102	127	151	206	222	286	473	100
3. milk & products	9	25	38	53	63	102	123	154	211	269	570	489	100
4. other food	36	55	69	81	97	104	127	142	167	177	219	364	100
5. total food	40	58	75	85	96	109	125	137	156	173	236	304	100
6. clothing	12	30	45	58	81	108	124	147	218	279	245	579	100
7. medical	2	21	26	44	63	86	107	119	195	281	402	907	100
8. others	25	38	45	56	71	89	105	138	160	243	286	744	100
9. non-food total	20	35	44	56	73	94	110	139	178	255	282	710	100
10. total (all items)	33	50	64	74	88	104	120	138	164	201	252	444	100
Index number of percentage of total expenditure													
11. foodgrains	158	141	139	129	120	110	104	94	80	70	61	48	100
12. sugar	53	78	77	95	94	98	106	110	126	110	114	106	100
13. milk & products	26	49	59	71	72	98	102	112	129	134	226	110	100
14. other food	108	109	108	108	110	100	106	103	102	88	87	82	100
15. food total	121	116	117	113	109	105	104	100	95	86	94	68	100
16. clothing	35	60	70	78	92	104	103	107	133	138	97	130	100
17. medical	7	42	40	59	72	83	89	86	120	140	160	204	100
18. others	76	75	71	75	81	86	87	100	98	121	113	168	100
19. non-food total	61	69	69	75	83	91	92	101	109	127	112	160	100
20. total	100	100	100	100	100	100	100	100	100	100	100	100	100
21. percentage of persons	14.01	17.31	11.74	9.14	11.55	8.21	7.04	5.53	5.17	4.85	2.59	2.86	100.00
22. cumulative percentage	14.01	31.32	43.06	52.20	63.75	71.96	79.00	84.53	89.70	94.55	71.14	100.00	—
23. size of households (persons)	112	111	118	104	100	95	93	91	84	93	68	70	100
24. no. of households	12.53	15.59	9.92	8.83	11.51	8.68	7.59	6.08	6.16	5.22	3.79	4.10	100
25. no. of persons	14.01	17.31	11.74	9.14	11.55	8.21	7.04	5.53	5.17	4.85	2.59	2.86	100
Index number of total expenditure													
26. on 'food' items	5.65	10.13	8.75	7.73	11.07	8.99	8.83	7.58	8.06	8.38	6.12	8.71	100
27. on 'non-food' items	2.89	5.99	5.15	5.07	8.47	7.75	7.76	7.69	9.22	12.37	7.32	20.32	100
28. on 'all' items	4.70	8.70	7.51	6.81	10.16	8.56	8.46	7.62	8.46	9.76	6.54	12.72	100
29. no. of sample households	260	369	240	217	305	244	197	207	193	187	138	197	2754

no. of sample villages = 1544, no. of sample urban blocks = 1220

Based on the National Sample Survey (NSS) : 10th Round, December 1955—May 1956, All-India (rural and urban) excluding Jammu & Kashmir.

9. The index-numbers given in Table A.1.1 lines 11 to 19, show the change in percentages of expenditures (as distinguished from actual expenditure) on different items. It is seen that the percentage expenditure on foodgrains and other items of food decreases with rising expenditure level. The percentage expenditures on sugar, milk and milk-products, and clothing, however, continue to increase as expenditure increases; and the relative increase is highest in the case of medical expense.

CONCENTRATION CURVES FOR CONSUMER EXPENDITURE

10. The question of equality or otherwise of the consumer expenditure can be studied in a different way in the form of what is called "concentration curves." Relevant data are given in Table A.1.2 and are also shown in a graphical form in curves B, C, and D, in the chart given on p. 76. Consider the right hand side of the chart. The horizontal scale at the bottom represents percentage of the population (beginning at the bottom with those having the lowest expenditure per person and finishing with those having the highest expenditure at the top). The vertical scale on the right hand side represents percentages of the total consumer expenditure, rising from zero to 100 percent at the top.

11. Consider the total expenditure, say, on all items and call this 100. If each person has an exactly equal share, then 5 percent of the population would have 5 percent of the total expenditure; 10 percent of the population would have 10 percent of the total expenditure and so on. It is possible to draw this as a graph. Corresponding to a point 5 percent on the horizontal scale the value of the ordinate would be 5 percent; for 10 percent on the horizontal scale the corresponding value on the vertical scale would be 10 percent. The graph is obviously given by the straight line joining the "zero" point on both scales with the "100"-point on both scales which is the straight line inclined at 45° to the axis. This line represents a completely egalitarian distribution without any concentration.

12. The actual distribution or apportionment of expenditure is, however, not equal; and it would be of interest to plot the graph of the observed expenditure. Relevant data are given in Table A.1.2, in which col. (1) shows the expenditure groups as in Tables A.1.0 and A.1.1. In this particular case the information is given separately for two sub-samples 1 and 2 based on different sets of sample villages and sample households. It may be mentioned here that the estimates given in Tables A.1.0 and A.1.1 were obtained by combining the two sub-samples 1 and 2.

13. The sample villages and the sample households in each sub-sample were selected in accordance with a design of sampling based on the theory of probability. Each sub-sample covers the whole of India (rural and urban), and each sub-sample would supply valid estimates for every item under study. The two sub-samples are completely interpenetrating; and the results based on the two samples have equal validity and are comparable. The object of using two (or more) interpenetrating sub-samples can be briefly explained. As each sub-sample gives a valid estimate, a direct comparison of the results based on two (or more) sub-samples immediately supplies information on the consistency of the estimates. If two (or more) sub-sample estimates are in good agreement, then it is possible to have confidence in the results. If the results diverge widely, then the results must be suspected. How this method works in practice would become clear as soon as the procedure for drawing the graphs is explained.

TABLE A.1.2. CUMULATIVE PERCENTAGE OF EXPENDITURE ON 'FOOD', 'NON-FOOD' AND 'ALL' ITEMS: ALL INDIA (RURAL AND URBAN) : 1955-56

per capita expenditure classes in Rs.	actual number of sample households		cumulative percentage of total number of persons		cumulative percentage of expenditure on					
					'food' items		'non-food' items		'all' items	
	S.S.1	S.S.2	S.S.1	S.S.2	S.S.1	S.S.2	S.S.1	S.S.2	S.S.1	S.S.2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
			....	....	....	......cumulative	percentage......	....	....	....
0—8	135	125	13.84	14.18	5.64	5.66	2.82	2.96	4.67	4.73
8—11	181	188	31.52	31.12	16.15	15.43	8.60	9.15	13.55	13.26
11—13	122	118	43.62	42.50	25.37	23.74	13.80	14.25	21.38	20.45
13—15	115	102	52.65	51.74	33.16	31.41	18.89	19.31	28.23	27.21
15—18	153	152	64.17	63.33	44.38	42.30	27.56	27.59	38.57	37.20
18—21	131	113	73.61	70.32	55.06	49.66	36.32	34.37	48.59	44.36
21—24	89	108	79.36	78.63	62.29	60.03	42.97	43.19	55.62	54.19
24—28	104	103	85.93	83.14	71.32	66.22	52.37	49.24	64.78	60.33
28—34	91	102	90.28	89.12	77.81	75.80	60.63	59.37	71.88	70.10
34—43	98	89	95.31	93.79	86.66	83.73	73.64	71.14	82.17	79.36
43—55	60	78	97.08	97.19	90.83	91.73	79.27	80.07	86.85	87.69
55 & above	98	99	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
all classes	1377	1377								

no. of sample villages : S.S.1 = 772, S.S.2 = 772. no. of sample blocks : S.S.1 = 610, S.S.2 = 610.

Based on National Sample Survey (NSS) : 10th Round, December 1955—May 1956, All India (rural and urban) excluding Jammu & Kashmir.

14. The number of sample households in the two sub-samples are given in cols. (2) and (3) to show the size of samples but are not used in drawing the concentration curves. Col. (4) gives the accumulated percentages of persons in sub-sample 1, similar to the figures given in line 22 of Table A.1.0. Percentages for sub-sample 2 are given in col. (5). The expenditure on, say, food items (or non-food items, or all items) is known for each expenditure group shown in col. (1) ; and the total expenditure on food items for all expenditure groups taken together is also known. It is, therefore, possible to express the expenditure incurred in any particular group as a percentage of the expenditure incurred by all the groups. In this way, each expenditure group, in each line of Table A.1.2 would have a figure giving its percentage share of the total expenditure on food (or non-food, or all) items. These percentages can then be added successively from the bottom up, and would supply the 'cumulative percentage of expenditure,' say, on food items shown in col. (6). For example, the expenditure group '0–Rs. 8' accounts for 5.64 percent of the total expenditure on food items. The two groups '0–Rs. 8' and 'Rs. 8–11' together account for 16.15 percent of the total expenditure on food items and so on.

15. The cumulative percentages of persons in col. (4) can be represented on the horizontal scale (at the bottom of the chart); and the corresponding cumulative percentage of expenditure can be represented on the vertical scale (on the right of the chart). The values given in the same line in cols. (4) and (6) are plotted on the chart and is shown by the graph for sub-sample 1 of concentration curve B for expenditure on food items. A similar graph can be drawn for sub-sample 2 on the basis of each pair of figures given in the same line of cols. (5) and (7). The two graphs for concentration curve B were drawn in this way. The same procedure was followed in the case of expenditure on non-food items by plotting the figures given in cols. (4) and (8) for sub-sample 1, and the figures in cols. (5) and (9) for sub-sample 2; these are shown in the two graphs of concentration curve C. For expenditure on all items, the figures in cols. (4) and (10) for sub-sample 1, and figures in cols. (5) and (11) for sub-sample 2 are plotted to give the two graphs of concentration curve D.

16. It is of interest to examine the three sets of double graphs of B, C, and D. Consider the two graphs of curve B, each of which gives an equally valid sample representation of the concentration curve for expenditure on food. They diverge to some extent but not too much. We may, therefore, infer that the 'true' concentration curve B for expenditure on food items lies somewhere in the neighbourhood of the pair of graphs for sub-samples 1 and 2. Also, the divergence between this pair of graphs supplies some idea of the margin of errors of sampling. In the same way, we can locate approximately (and always subject to some margin of errors of sampling) the position of concentration curve C for expenditure on non-food items; and the position of concentration curve D for expenditure on all items of consumption.

17. The divergence between the concentration curve B, on one hand, and the concentration curve D, on the other hand, can be directly seen on the chart to be much greater than the divergence between the pair of graphs for the two sub-samples of curve B or of curve D. From this we may infer that the two concentration curves B and D are significantly different in the statistical sense. Looking at all the three pairs of graphs it can be inferred that the three concentration curves B, C and

TABLE A.2.0. PERCENTAGE OF HOUSEHOLDS (EXCLUDING THOSE OF SIZE LESS THAN 0.10 ACRE) BELOW SPECIFIED SIZE OF OWNERSHIP HOLDING AND THE CORRESPONDING PERCENTAGE OF AREA OWNED : ALL INDIA (RURAL) : 1953-54

	specified size of ownership holdings	number of sample households			cumulative percentage of					
					total number of households			area owned		
		Central	State I	State II	Central	State I	State II	Central	State I	State II
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1.	0.10–0.50 acre	1107	1210	1043	9.49	10.94	9.47	0.37	0.40	0.33
2.	upto 1.00 ,,	1099	1095	1075	18.95	20.58	18.82	1.29	1.37	1.23
3.	,, 1.50 ,,	1060	996	1012	27.38	28.90	27.34	2.70	2.80	2.61
4.	,, 2.50 ,,	1656	1578	1596	40.21	41.33	40.75	6.17	6.20	6.10
5.	,, 5.00 ,,	2856	2805	2622	61.24	61.97	60.68	16.53	16.65	15.54
6.	,, 7.50 ,,	1825	1848	1722	73.73	73.51	72.60	26.99	26.55	25.20
7.	,, 10.00 ,,	1164	1265	1241	80.77	80.60	80.05	35.34	35.17	33.73
8.	,, 15.00 ,,	2311	2495	2486	88.25	88.69	87.71	47.93	48.93	46.06
9.	,, 20.00 ,,	1368	1337	1337	92.42	92.56	91.78	57.83	58.28	55.36
10.	,, 25.00 ,,	1038	964	1021	94.56	94.80	94.11	64.38	65.23	62.18
11.	,, 30.00 ,,	675	629	740	96.13	96.15	95.76	70.29	70.38	68.18
12.	,, 40.00 ,,	757	693	847	97.77	97.77	97.56	77.98	78.19	76.32
13.	,, 50.00 ,,	396	392	414	98.61	98.64	98.48	83.10	83.58	81.39
14.	,, 75.00 ,,	413	395	464	99.48	99.44	99.44	90.16	90.37	89.48
15.	,, 100.00 ,,	127	134	136	99.73	99.71	99.70	92.98	93.64	92.40
16.	,, 250.00 ,,	131	134	128	99.97	99.99	99.97	97.60	99.10	97.69
17.	,, 500.00 ,,	14	6	18	99.99	100.00	100.00	98.62	99.68	99.01
18.	above 500.00 ,,	6	1	4	100.00	100.00	100.00	100.00	100.00	100.00
19.	total	18,003	17,977	17,906						
20.	sample villages	1,410	1,422	1,383						

Based on the National Sample Survey : 8th Round, July 1954—March 1955. Information for the Central sample was collected by an agency of the Government of India and for the State samples by the respective State Governments. The Central, State-I, and State-II samples are each based on four interpenetrating samples; and all 12 samples also form, of course, an interpenetrating net work of samples covering the rural area.

D are all different. It has been possible to make this inference in this very simple manner, because we have been able to use a pair of graphs (based respectively on sub-samples 1 and 2) for each of the three concentration curves B, C and D.[1]

18. We may now examine the concentration curve or the figures given in Table A.1.2 to understand the implications. For sub-sample 1, the lowest group '0 to Rs. 8' has a percentage share of population of 13.84 percent, in col. (4); but its share of expenditure on food items is only 5.64 percent, in col. (6); and the share of non-food expenditure is still less and only 2.82 percent, in col. (8); the share of expenditure on all items is 4.67 percent, in col. (10), and naturally falls between the other two. The position can be ascertained for other groups or for any percentage of the population directly from the graph. It would be seen the lower the concentration curve at the beginning, the lower will be the share of the poorer households.

19. It will be noticed that expenditure on food is more egalitarian than the expenditure on all items, while expenditure on non-food items deviates most from equality. This can be seen clearly from the curves B, C and D on the chart. The further a concentration curve lies towards the lower right-hand corner of the chart, the greater is its deviation from the egalitarian distribution shown by the straight line inclined at 45° to the axis. The more intense the concentration, the greater would be the advantage enjoyed by a few rich men over a very large number of poor people. Concentration curves thus give a simple visual picture of the non-equality of the distribution.

CONCENTRATION CURVE FOR LAND HOLDINGS

20. It is, of course, possible to draw concentration curves for distributions other than that of expenditure. The case of holdings of land by households is of great interest. Some relevant data are given in Table A. 2.0 based on a special survey of land-holdings carried out in the NSS : 8th round : July 1954–March 1955 covering the whole rural area of India. In this survey 12 interpenetrating sub-samples were used, information for 4 of which was collected by the NSS (which is a Central agency) and information for the remaining 8 sub-samples was collected by respective State Governments in the form of two groups of samples called State-I and State-II, each consisting of 4 sub-samples. The information is given in Table A. 2.0 separately for 3 groups of samples : Central, State-I and State-II.

21. Information was obtained about the size in acres of the total holdings of land owned by each household. The households were then classified in groups by size of holding, such as, owning '0.10 to 0.50 acre;' owning '0.50 to 1.00 acre' and so on. The number of households in each group and also the total area of land held by each group were then calculated; and these were expressed as percentage of the total number of households and of the total area of land owned by all households. The next step was adding up the percentages by successive groups to give the 'cumulative percentage of

---

[1] When a particular number or quantity (as distinguished from a curve, as in the present case) is estimated, the margin of error can be calculated with considerable accuracy on results based on two or more interpenetrating samples. The rule is extremely simple: for a set of 2, 3, 4, 5, 6, ... k interpenetrating samples the probability would be respectively

$$1/2,\ 3/4,\ 7/8,\ 15/16,\ 31/32\ \ldots.\ (2^{k-1}-1)/2^{k-1}$$

and the 'true' value of the parameter lies between the two extreme estimates based on the interpenetrating samples. This method is being used in the Indian Statistical Institute since 1937.

household' and the 'cumulative percentage of total area of land owned by households.' These figures are given for the Central sample in cols. (5) and (8) respectively of Table A.2.0; for State-I sample in cols. (6) and (9) respectively; and for State-II sample in cols. (7) and (10) respectively. These three sets of values were then drawn on the chart in the form of the three graphs of concentration curve E, at the extreme lower right-hand corner of the chart.

22. The number of sample households is given in cols. (2), (3) and (4) which show that the size of the sample was much larger in the case of the survey of land-holdings compared to the survey of household expenditure. Each of the Central, State-I and State-II samples had about 18,000 sample households compared to only 1377 sample households in sub-samples 1 and 2 in the case of household expenditure. It is of interest to note that the three graphs of curve E lie very close to one another (and are almost indistinguishable), no doubt, due to the large size of the samples. Comparing the graphs of curve E with the graphs of B, C and D, it is seen that the margin of errors of sampling is much smaller in the case of curve E. The size of samples in the case of land-holdings would seem to be quite adequate.

23. It should be mentioned that in the present case only those households have been included which own more than 0.10 acre of land in order to exclude landless labour and households who practically do not own any land. The most important point to be noted is that the concentration of land-holdings is far greater than the concentration of expenditure on non-food items of consumption. The distribution of land-holdings is thus far more unequal than the distribution of consumer expenditure. The curves B, C, D and E give a clear idea of the inequality in the level of living and in the ownership of land in India at the present time.

EMPLOYMENT AND UNEMPLOYMENT

24. Table A.3.1 shows the position regarding the labour force, employment, and unemployment based on the 9th Round of the National Sample Survey (NSS) : May–August 1955. Estimated numbers in million are given for each category in cols. (2), (3) and (4) for rural and urban areas, and all-India respectively. The corresponding percentage of the total are shown in cols. (5), (6) and (7). The first portion A gives the basic information. Consider the all-India figures given in col. (4) and corresponding percentages in col. (7). It is seen from line 1 that the total estimated population (on the basis of the National Sample Survey) was 382.4 million in 1955, and from line 3 that the labour force consisted of 162.6 million persons [or 42.5 percent of the population, as shown in col. (7), line 3] who were at work or seeking work at the time of the survey. Within the labour-force itself 160.4 million were gainfully occupied (line 5) and 2.24 million were "unemployed" (line 4) in accordance with the concept of "unemployment" used in industrially advanced countries.

25. For various reasons, it is known that the concept of unemployment used in the advanced countries is not suitable in the case of under-developed countries in which most of the people work in household enterprises and agriculture. They are never "unemployed" in the technical sense, but may often have to sit idle for lack of gainful work. The figures for "unemployed" given in line 4 are not, therefore, adequate. A better picture can be obtained, if consideration is given to the amount of time a person

is available for work or is actually engaged in gainful work. This has been sought to be done in Part B of the table. Three different approaches were used.

26. Information was collected on how many more hours a person was available for work per week and the replies were classified in groups of "available from 1 to 8 hours per week," "from 8 to 14 hours per week," etc., (that is, by units of 1 hour per day per week), shown in lines 6 to 13. Those who are available for work for more than 29 hours per week (that is, who can accept work for more than 4 hours per day for 7 days) are either practically unemployed or are sitting idle for half the week or more. Their number would come to 11.7 million [col. (4), line 9]. This is perhaps a lower limit of effective unemployment.

27. Information about hours of work per week is given (in an accumulated form) in lines 14 to 20. It would be seen that the number of persons working less than 8 hours per week was 20.9 million and their position was practically the same as that of the unemployed. The number of persons working less than 29 hours per week or 4 hours per day was as large as 44.8 million [col. (4), line 16].

28. A third approach was to enquire how many days a person was engaged in gainful work during 30 days preceding the day of interview. The information is given in the form of accumulated totals in lines 21 to 26. The number of persons working less than 10 days out of 30 days, or less than one-third of the month, was 39.2 million [col. (4), line 22]. This is broadly consistent with the figure 44.8 million given in the previous paragraph.

29. There is, however, a big gap between estimates of the number of unemployed persons based on B.(1) the hours per week available for additional work and B.(2) the hours of work per week. A special enquiry carried out in the NSS: 7th Round, October 1953–March 1954 had given some interesting information on this point. The number of persons working part-time was obtained as 47.2 million corresponding roughly to 44.8 million persons who worked 28 hours or less per week as given in col. (4), line 16 of Table A.3.1. The enquiry in the 7th Round brought out that among them 10.4 million (22 percent of the total) were working part-time partly on account of ill-health of the worker and partly because of illness in his household. Also, 23.4 million persons (or roughly half) were working a part of the time for reasons of an economic nature, such as lack of demand for their labour, lack of tools and raw materials, slack and off season, etc.; and 13.4 million persons gave other reasons. It would seem from the above discussion that at least 23 million persons, who were working part-time for reasons of an economic nature, can be considered as severely "underemployed."

30. It is clear that the volume of underemployment is very great indeed in India. It is, however, difficult to set up precise definitions and standards for unemployment and underemployment especially in rural areas and in household and small scale enterprises and in the case of self-employed persons. The actual number would depend entirely on where the line is to be drawn between employment and underemployment.

31. Table 3.2 gives information on the distribution of gainfully employed persons by industries and by economic sectors (public, large scale, private household, etc.). From col. (9) it is seen that out of a total number of gainfully employed persons of 160.38 million, 119.76 or nearly 120 million work in agriculture, forestry, livestock, etc. The total number engaged in manufacturing industries is only 17.54 million, out

TABLE A.3.1. STATUS OF ECONOMIC ACTIVITY AND GAINFUL EMPLOYMENT : RURAL, URBAN & ALL INDIA

		estimated number of persons in million			percentage of total		
		rural	urban	all-India	rural	urban	all-India
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	A. Distribution of population by status of economic activity[1]						
1.	total population	316.1	66.3	382.4	100.0	100.0	100.0
2.	outside labour force	176.4	43.4	219.8	55.8	65.6	57.5
3.	in labour force	139.7	22.9	162.6	44.2	34.5	42.5
4.	unemployed	0.92	1.32	2.24	0.29	1.99	0.59
5.	gainfully employed	138.8	21.6	160.4	43.9	32.6	41.9
	B. Distribution of gainfully employed persons by period of work						
(1)	*hours available for additional work per week*[2]						
6.	above 70 hours	0.38	0.04	0.42	0.28	0.20	0.26
7.	,, 56 ,,	1.2	0.11	1.3	0.85	0.51	0.81
8.	,, 42 ,,	6.1	0.7	6.8	4.4	3.3	4.2
9.	,, 28 ,,	10.4	1.3	11.7	7.5	6.2	7.3
10.	,, 14 ,,	17.6	2.5	20.1	12.7	11.6	12.5
11.	,, 7 ,,	22.1	3.2	25.3	15.9	15.0	15.8
12.	,, 1 hour	24.3	3.4	27.7	17.5	15.7	17.3
13.	nil	138.8	21.6	160.4	100.0	100.0	100.0
(2)	*hours of work per week*[2]						
14.	7 hours or less	18.9	2.0	20.9	13.6	9.2	13.0
15.	14 ,,	24.3	2.6	26.9	17.5	12.1	16.8
16.	28 ,,	40.2	4.6	44.8	29.0	21.5	27.9
17.	42 ,,	64.3	8.5	72.8	46.3	39.5	45.4
18.	56 ,,	109.7	17.2	126.9	79.0	79.4	79.1
19.	70 ,,	133.4	20.3	153.7	96.1	93.9	95.8
20.	all	138.8	21.6	160.4	100.0	100.0	100.0
(3)	*days of work per 30 days*[3]						
21.	less than 5 days	28.2	1.6	29.8	30.2	7.2	18.6
22.	,, ,, 10 ,,	37.2	2.1	39.2	26.7	9.6	24.4
23.	,, ,, 15 ,,	50.1	3.2	53.3	36.1	14.6	33.2
24.	,, ,, 20 ,,	62.5	4.6	67.1	45.0	21.1	41.8
25.	,, ,, 25 ,,	78.3	8.2	86.5	56.4	38.0	53.9
26.	,, ,, 30 ,,	138.8	21.6	160.4	100.0	100.0	100.0

---

[1] Based on the National Sample Survey (NSS): 9th Round, May–August 1955; growth of population calculated on the Census 1941–1951 rate of 1.32 percent per annum.

[2] Percentages in cols. (5) to (7) and lines (1) to (20) based on NSS : 9th Round, May–August 1955, sub-samples 1 and 2 only, consisting of 8,037 rural and 16,703 urban sample households.

[3] Percentages in cols. (5) to (7) and lines (21) to (26) based on NSS : 5th and 6th Rounds, December 1952–August 1953, consisting of 12,976 rural and 5,670 urban sample households.

of which also only 3.47 million are engaged in modern large scale industries. Trade and commerce with 5.84 million and service with 5.13 million come next in order of importance.

32. From col. (2), bottom line, it is seen that 7.26 million persons are employed in the public sector, that is, under the Central and State Governments, Municipalities, and other public authorities; and from the total of col. (3) that 5.65 million are employed in private large scale enterprises. These two taken together, 12.91 million persons have employment of the type familiar in industrially advanced countries; and in their case it is probably possible to specify fairly well whether a person is employed or unemployed in the sense in which "unemployment" is defined in advanced countries.

33. The private small scale sector has 146.61 million [col. (7), bottom line], out of which 115.23 million are household labour (col. 4). They may be fully or only partly engaged in work but can almost never be "unemployed" in the technical sense. The structure of employment clearly shows the backwardness of the Indian economy.

TABLE A.3.2. GAINFULLY EMPLOYED PERSONS BY INDUSTRIES : ALL INDIA : 1955

industry	public	private large-scale	private small-scale				not recorded	total
			house-hold labour	hired labour	domes-tic service	sub-total (4)+(5) +(6)		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
				million persons				
1. agriculture, forestry, livestock, fishing and hunting	0.41	0.93	93.23	24.76	0.00	117.99	0.51	119.84
2. mining and quarrying	0.04	0.39	0.01	0.08	—	0.09	0.00	0.52
3. manufacturing	0.44	3.47	11.38	2.10	0.00	13.48	0.16	17.55
4. construction	0.26	0.22	0.66	0.85	0.00	1.51	0.01	2.00
5. electricity, gas, water, sanitation	0.16	0.01	0.09	0.00	—	0.09	0.01	0.27
6. trade and commerce	0.10	0.22	4.78	0.72	0.00	5.50	0.02	5.84
7. service	2.65	0.14	1.07	0.49	0.77	2.33	0.01	5.13
8. transport and communication	1.47	0.14	1.03	0.24	0.00	1.27	0.03	2.91
9. not specified above and not recorded	1.73	0.13	2.98	0.67	0.70	4.35	0.11	6.32
10. total	7.26	5.65	115.23	29.91	1.47	146.61	0.86	160.38

Based on the National Sample Survey (NSS) : 9th Round : May—August 1955, sub-samples 1 and 2 only consisting of 8,027 rural and 16,703 urban households.

34. Table A.3.3 shows the incidence of "unemployment" among educated persons defined as matriculates or having higher educational qualifications. The concept of "unemployment" used is the same as that used in the advanced countries and is known to be much too restricted and hence unsuitable for underdeveloped countries. However,

using this strict definition of unemployment, it is of interest to observe [from col. (7), line 5] that there were about 0.47 million or 4.7 lakhs unemployed in 1955 out of a total of 5 million or 50 lakhs of educated persons altogether. Taking into consideration the facts that about 4 lakhs of candidates matriculated in 1955 and that the number is steadily increasing, it can be easily imagined how very difficult the problem of unemployment among educated persons would become in future, unless there is rapid economic development.

TABLE A.3.3. DISTRIBUTION OF TOTAL POPULATION AND UNEMPLOYED PERSONS BY EDUCATION CLASS : RURAL, URBAN & ALL INDIA : 1955

education class	rural		urban		all-India		percentage of unemployed		
	total	un-employed	total	un-employed	total	un-employed	rural	urban	all-India
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	.... million persons						....percentage....		
1. below matriculate	314.70	0.76	62.70	1.01	377.40	1.77	0.24	1.60	0.47
2. matriculate	1.10		2.40		3.50				
3. intermediate	0.19		0.65		0.84				
4. graduate and above	0.13		0.57		0.70				
5. sub-total (2—4)	1.42	0.16	3.62	0.31	5.04	0.47	11.42	8.59	9.40
6. total	316.12	0.92	66.32	1.32	382.44	2.24	0.29	1.99	0.59

Based on the National Sample Survey : 9th Round, May–August 1955, sub-samples 1 and 2 only, consisting of 8,037 rural and 16,703 urban sample households. Growth of population calculated on the basis of the Census 1941–51 rate of 1.32 percent per annum.

GENERAL AND PROFESSIONAL EDUCATION

35. Table A.3.4 shows the number of persons passing different examinations in India each year from 1950 to 1955 based on information collected by the Planning Commission. There is no doubt that general education is increasing quite fast. In most cases numbers have more than doubled in five years since 1950. The number of matriculates was over 4 lakhs in 1955 which represented about 1050 matriculates per million of the population. The number of first graduates was about 57,000 or about 150 per million which is comparable with an out-turn of about 20,000 graduates in the UK with a population of 51 million or 392 per million. A comparison can also be made on the basis of literate persons. The number of literates in India in 1955 was about 63.2 million (all ages below 5 are omitted in calculations relating to literacy). The number of matriculates in 1955 was thus about 5300 per million literates; and the number of first graduates was about 900 per million literates. In the UK the number of persons of age 5 years or more was about 44 million and this would be also the number of literates (because of universal literacy) and is comparable to 63 million literates in India. On the basis of an out-turn of about 20,000 graduates per year in the UK, the number would work out at about 450 graduates per million literates. The proportion of graduates in India in relation to literates (900) would thus seem to be double that of the UK (450). This supplies some interesting information on social differentials in education.

TABLE A.3.4. GENERAL AND PROFESSIONAL EDUCATION : ALL INDIA 1950–1955

qualification	number of persons qualifying in 1950	1951	1952	1953	1954	1955	index-number: 1951 = 100
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
General education							
1. Matriculation	1,89,184	2,41,143	2,61,059	3,34,760	3,97,005	4,00,014	166
2. Intermediate	59,283	72,685	77,836	89,021	1,04,851	1,26,476	174
.1 Arts	41,652	47,013	52,373	59,703	71,640	86,741	185
.2 Science	17,631	25,672	25,463	29,318	33,211	39,735	155
3. First degree	28,745	32,238	36,136	40,033	50,178	57,051	177
.1 Arts	19,212	21,251	24,965	27,491	35,773	40,444	190
.2 Science	9,533	10,987	11,171	12,542	14,405	16,607	151
4. Master's degree	5,581	7,138	7,734	7,855	9,777	11,013	154
.1 Arts	4,503	5,729	6,054	6,161	7,709	8,802	154
.2 Science	1,078	1,409	1,680	1,694	2,068	2,211	157
5. Doctorate	115	146	164	123	280	324	222
.1 D.Phil.(junior)	56	136	159	115	269	n.a.	(198)
.2 D.Sc.(senior)	59	10	5	8	11	n.a.	(110)
6. total	2,82,908	3,53,350	3,82,929	4,71,792	5,62,091	5,94,878	168
Engineering and technology							
7. Certificate	566	14,189	13,604	17,418	15,718	18,588	131
8. Diploma	784	1,547	2,871	2,575	3,148	3,774	244
9. First degree	1,923	2,189	2,510	3,046	3,520	3,583	164
10. Master's degree	50	66	82	69	211	144	218
11. total	3,323	17,991	19,067	23,108	22,597	26,089	145
Medical science							
12. Diploma	50	117	120	267	—	—	—
13. First degree	2,017	1,624	1,985	2,146	3,131	—	(193)
14. Master's degree	19	256	261	350	693	—	(271)
15. Doctorate	52	54	55	63	48	—	(89)
16. total	2,138	2,051	2,421	2,826	3,872	—	(200)
Agriculture							
17. Certificates	459	783	1,120	982	999	1,510	193
18. Diploma	585	661	352	691	563	778	118
19. First degree	1,067	1,118	984	914	1,004	971	87
20. Master's degree	146	151	209	190	198	193	128
21. total	2,257	2,713	2,665	2,777	2,764	3,452	127
Veterinary science							
22. Diploma	62	34	37	—	—	—	—
23. First degree	112	242	237	252	298	300	124
24. Master's degree	—	—	—	—	135	147	—
25. total	174	276	274	252	433	447	124

Notes : (1) Index-numbers within brackets are for year 1954.

(2) Source of information : Division of Perspective Planning, Planning Commission.

CORRIGENDA

Some observations on Input-output Analysis : By Oskar Lange, *Sankhyā* **17**, 305-36.

A statement contained in my paper "Some Observations on Input-Output Analysis" which appeared in *Sankhyā*, Vol. 17 Part 4 (February 1957) needs correction and further elucidation. This is the contention on pp. 331-332 that the rate of increase in national income is equal to the rate of increase of gross national product. This is so only when the change taking place in replacement is neglected. Taking account of the latter, we obtain the following result.

Using the notation in my paper national income in unit period t is

$$\sum_i x_i(t) = \sum_i X_i(t) - \sum_i \sum_j a'_{ij} X_j(t).$$

The double sum on the right hand side represents that part of gross national product which is used for replacement of the means of production used up during the unit period, i.e., amortisation. Let us denote the rate of replacement (rate of amortisation) during unit period t by

$$\sigma(t) = \frac{\sum_i \sum_j a'_{ij} X_j(t)}{\sum_i X_i(t)},$$

we then can write

$$\sum_i x_i(t) = \sum_i X_i(t)[1-\sigma(t)].$$

Denoting by $r(t)$ the rate of increase of national income and by $R(t)$ the rate of increase of gross national product we obtain

$$\frac{\sum_i x_i(t+1)}{\sum_i x_i(t)} = 1+r(t) = \frac{\sum_i X_i(t+1)[1-\sigma(t+1)]}{\sum_i X_i(t)[1-\sigma(t)]}$$

whence the relation

$$1+r(t) = [1+R(t)] \frac{1-\sigma(t+1)}{1-\sigma(t)} \qquad \dots \quad (1)$$

between the rate of increase of national income $r(t)$ and the rate of increase of gross national product $R(t)$.

In the relation obtained $1+r(t)$ is the growth coefficient of national income, $1+R(t)$ is the growth coefficient of gross national product. $1-\sigma(t)$ and $1-\sigma(t+1)$ represent the fraction of gross national product in the respective unit periods t and $t+1$ which do not serve for replacement; it may be called the net output ratio. Thus relation(1) states that the growth coefficient of national income is equal to the growth coefficient of gross national product multiplied by an expression indicating the change in the net output ratio.

National income in unit period t_s is related to national income in the initial unit period t_0 $(t_s > t_0)$ by formulae analogous to (7.10) and (7.11) in my paper, namely,

$$\sum_i x_i(t_s) = \prod_{t=t_0}^{t_s} [1+r(t)] \,.\, \sum_i X_i(t) \qquad \dots \quad (2)$$

and, in case when $r(t) = r = \text{const}.$

$$\sum_i x_i(t_s) = (1+r)^{t_s - t_0} \sum_i x_i(t_0).$$

In view of (7.9) in my paper and (2) in this note the expression (7.14) can also be written in the more explicit form

$$\sum_i x_i(t_s) = \prod_{t=t_0}^{t_s} \left\{ [1+\alpha(t)\beta(t)] \frac{1-\sigma(t+1)}{1-\sigma(t)} \right\} . \sum_i x_i(t). \qquad \dots \quad (3)$$

These results have to be substituted for the formulae (7.13), (7.14) and (7.15) of my paper.

CORRIGENDA

"A note on Distance between Two Populations" By S. R. Adke, *Sankhyā* **19,** 195-200.

1) Relation 3.5 should read

$$\sqrt{(n-1)\,J} = t = \sqrt{-(n-1)\,8 \log\left(1-\tfrac{1}{2} I_2\right)}.$$

2) Line 2, on page 198, should read

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left[-\{\Sigma\Sigma a_{ij}x_ix_j + 2\Sigma b_ix_i + c\}\right] dx_1 \dots dx_p$$

$$= \frac{\pi^{p/2}}{|a_{ij}|^{\frac{1}{2}}} \exp\left[-\frac{|\Delta|}{|A|}\right].$$

However, the above corrections do not alter the results obtained later in the paper.

INDIAN STATISTICAL INSTITUTE

TWENTY-FIFTH ANNUAL REPORT : April 1956—March 1957

PART 1 : CONSTITUTION AND ACTIVITIES OF THE INSTITUTE

1. The Indian Statistical Institute completed the twenty-fifth year of its existence on 17 December of the year under review (April 1956 to March 1957). Though work on statistics on scientific lines had started much earlier under the guidance of Professor Mahalanobis in what was known as the Statistical Laboratory, the Institute as such came into existence by a resolution passed at a public meeting held on 17 December 1931 under the Chairmanship of the late Sir R. N. Mookerjee who was elected the first President of the Institute and held that office for 5 years (1931–36). The Institute was registered as a non-profit learned society under Act XXI of 1860 in April 1932.

2. Starting with a single part-time computer and a total current expenditure of Rs.2 38 in the first year, the Institute has passed through alternating stages of acute financial difficulty and relative prosperity in reaching its present dimensions when it has a staff of over 1750 paid workers and a current expenditure budget of about Rs. 72 lakhs. During the first few years of its existence the Institute was mostly engaged in analytical studies including design of experiments. Gradually it began to undertake small-scale enquiries (indebtedness of agriculturists, economic conditions of handloom weavers, consumer preferences, etc.) on behalf of Government departments and private concerns. From 1936 the Institute became increasingly involved in crop estimating and land utilization surveys undertaken on behalf of the Indian Central Jute Committee, the Government of Bengal and the Government of Bihar and developed the technique of large-scale sample surveys for this purpose. Since 1950 the Institute has taken a leading part in planning, organizing and carrying out the National Sample Survey at the desire of the Government of India; since 1954 it is actively helping in the work of planning for national development. A short history of the origin and development of the Institute is given in Appendix 1 of this report.

3. *Constitution :* The supreme control of the Institute is vested in the general body of members of the Institute, which includes ordinary, life and honorary members, honorary fellows and patrons. It is this body which has the authority to make and amend the rules of the Institute. The Annual Report and Audited Accounts are presented at the Annual General Meetings; other General Meetings are held whenever necessary.

4. *Management :* The management of the Institute is vested in a Council consisting of the President, Chairman, two or more Vice-Chairmen, Treasurer, Secretary and other office-bearers, and 20 members, elected by the members of the Institute, and ex-officio representatives of regional branches and co-opted members. Subject to overall coordination by the Council, the management of the Research and Training School is vested in a

Governing Body with elected representatives of the Council, of the Government of India, and of a number of public bodies. There are two finance committees, one of the Governing Body and the other of the Council; an *ad hoc* committee for allocation of expenditure on which the Government of India is represented; a journal committee and an examinations committee. There are several other committees, executive and technical, for carrying out the day-to-day administration and for coordination of the work of different departments.

5. *Audit :* The Institute has a chartered accountant as internal auditor, while a firm of chartered accountants, qualified to audit accounts under the Companies Act, is appointed by the General Body of members at the Annual General Meeting as external auditor.

Activities

6. The Institute's activities which cover a wide field can be grouped into the following sectors :

1) *Society-type Activities :* The Institute has, since its establishment, functioned as a learned society, with *Sankhyā : The Indian Journal of Statistics* as its official organ. It has society-type branches in Aligarh, Bangalore, Bombay, Delhi, Madras and Poona, each branch having its own managing organization. The Institute also promotes a two-way flow of scientific visits for maintaining contacts with scientific institutions and workers in different parts of India and in other countries of the world.

2) *Research and Training School (RTS) :* The Institute maintains a Research and Training School for post-M.Sc. training in professional statistics, and offers facilities for advanced study and research in both theoretical and applied subjects. The RTS has well-equipped departments for theoretical and applied statistics, biometry, psychometry, demography, anthropology and applied economics. The School also organizes the training of statistical computers at different levels and conducts several other special courses of training in collaboration with the Central Statistical Organization of the Government of India.

3) *Examinations :* Since 1938 the Institute has been holding all-India examinations in a number of centres in the Indian Union for the award of Statistician's Diploma and Computer's Certificates and also Certificates of Proficiency in Field-Work.

4) *Projects :* The Institute has for many years been actively carrying out large-scale sample surveys for the collection of information on agricultural crops and social and economic conditions of people on behalf of different governments and other parties with the help of contract grants. The Institute has been handling since 1950 the big project of the design and technical work of the Indian National Sample Survey financed by the Government of India. A very large portion of the Institute staff is at present engaged in this project.

5) *Electronic Computer :* The Institute has an Electronic Computation Laboratory, which maintains an electronic digital computer HEC-2M and renders computation service

to different departments of the Institute and other scientific institutions in India. The Laboratory is also engaged, in collaboration with the Workshop, in developing devices and equipment for electronic computers.

6) *Development Workshop :* The Institute has a well-equipped Workshop which, in addition to the maintenance and servicing of calculating machines and instruments and equipment of the Institute, assists the Electronic Computer Laboratory in the maintenance and development of electronic computers. The Workshop is carrying out developmental work for the manufacture of desk calculating machines and scientific instruments. It has recently received many items of precision tools and equipment from the USSR through the United Nations.

7) *Statistical Quality Control (SQC) :* The Institute maintains three whole-time Statistical Quality Control (SQC) Units, at Bangalore, Bombay and Calcutta to give service to industrial concerns on a fee basis and also to promote the use of SQC generally. General guidance is given in this matter by a Policy Advisory Committee set up by the Government of India.

8) *Operational Research Relating to Planning :* The Institute is engaged in Operational Research relating to planning for national development since 1954 when studies in this connexion were inaugurated by the Prime Minister Sri Jawaharlal Nehru. The first phase of the work led to the preparation of the "Draft Plan Frame" which was submitted to Government in 1955 and was accepted as the basis for the formulation of the Second Five Year Plan. Since then one unit has been engaged in economic analysis and studies relating to planning, while another unit posted in Delhi is working on problems of scientific and technical man-power and perspective planning in collaboration with the Planning Commission. An experimental unit for the study of the economics of cottage industries, called *Kalyanasri*,* was started in Calcutta in 1955. Another unit, the Industrial Management Research Unit for Planning (IMRUP), was also started in 1955 with its headquarters at Bangalore for research and studies in industrial management. Another unit was started in 1956 for research in regional planning with emphasis on geographical surveys with headquarters at Bangalore. The work on planning is being done in collaboration with the Central Statistical Organization, the Planning Commission and relevant Ministries of the Government of India.

9) *General Services :* The Institute has a central library consisting of over 60,000 volumes and more than 1,200 journals and periodicals. There is a Translation Unit attached to the Library which is engaged in translating documents of interest for purposes of planning from the Chinese, the Japanese and the Russian languages. There is also a well-equipped photographic section. The Institute has a Machine Tabulation Unit with punched card equipment of both Hollerith and Powers-Samas types consisting of about 20 sorters, 18 tabulators and necessary complements of multipliers, collators, reproducers and gang-punching equipment in Calcutta and a smaller unit at Giridih and Delhi. There is an Estate Department for the construction, repair and maintenance of the Institute's buildings, roads, grounds, furniture and other equipment.

---

* The spelling adopted in this Report is 'Sri' instead of 'Shri or Shree' in accordance with the existing international convention for transliteration from the Devanagri (or Sanskrit) into the Roman script.

10) *Social Services :* The Institute maintains hostels for students, a canteen for supplying subsidized meals and refreshments, a circulating library, medical units including a health home at Giridih, a night school, an adult literacy centre, an educational aid unit, and a workers' club for sports, recreation and cultural activities. The Institute also publishes a house-journal, "Samvadadhvam." It provides for a good deal of "in-service" training. There is also a workers' Cooperative Credit Society. The Institute maintains guest-houses for visitors and a transport unit whose main function is to provide transport for the conveyance of workers between the Institute premises at Baranagar and the city.

7. In addition to the work done under the direct control of the Institute, the following associated organizations located in or near the Institute work in close cooperation with the Institute :

1) *International Statistical Education Centre (ISEC) :* Established in 1950 under the auspices of the UNESCO it is run under the joint direction of the International Statistical Institute and the Indian Statistical Institute with the support of the Government of India.

2) *Statistical Publishing Society :* This was established in 1935 on the initiative of the Institute as a non-profit association and registered under Act XXI of 1860 for the publication of Sankhyā : The Indian Journal of Statistics. The Society maintains the Eka Press which is adequately equipped for undertaking scientific and technical work.

3) *Indian Calculating Machine and Scientific Instrument Research Society :* This was established in 1943, on the initiative of the Institute, as a non-profit society registered in 1943 under Act XXI of 1860 with the object of promoting research, study, production and use in India of calculating machines, statistical, mathematical, scientific and engineering instruments and apparatus and appliances. This Society is now working in association with the Institute Workshop.

8. The main premises of the Institute are at present located on the Institute's own land of 4.78 acres at 203 Barrackpore Trunk Road in a multi-storeyed building with a floor-space of over 62,000 sq. ft. and about 21,250 sq. feet of ground-floor space in sheds. In addition, about 60,000 sq. feet of floor-space are occupied by the Institute at 202 Barrackpore Trunk Road which has been placed at the disposal of the Institute by the Government of In ia. The Institute also uses about 50,000 sq. feet of space in rented premises in Baranagar and Calcutta. The Institute has about 35 acres of land and a working branch with a staff of about 130 persons at Giridih. It also maintains offices at Bangalore, Bombay and Delhi.

9. *Twenty-fifth Anniversary :* The 25th Anniversary of the Institute was celebrated in December 1956 and January 1957. Members of the staff and their friends and many foreign and Indian Scientists attended the two general meetings held on 17 December 1956 and on 22 December 1956 in the mango grove on the Institute's grounds. The most important feature of the celebrations was a programme of about 60 lectures and seminars by leading statisticians and economists from different parts of India and abroad, spread over nearly three weeks. There was also a crowded programme of entertainments, such as a musical soiree, staging of Bengali plays, a folk play and cinema shows. A detailed account is given in Appendix 2.

TWENTY-FIFTH ANNUAL REPORT : 1956–57

PART 2 : YEAR UNDER REVIEW

1. *Planning for National Development :* When the Institute undertook its first sample survey way back in the early thirties, the first step was taken in a direction which was ultimately to result in the Institute undertaking planning for national development as its principal objective. It is in terms of this objective that a brief survey can best be made of the work of the Institute during the year under report (April 1956–March 1957).

2. In the Planning Division of the Institute, while an attempt was made to prepare an annual plan, the main emphasis was on long-range or perspective planning. Particular mention should be made in this connexion of Professor N. Kaldor's work on taxation undertaken under the auspices of the Institute which has since been largely drawn upon by the Union Government in framing its latest budget.

3. *Industrial Management Research :* The Industrial Management Research Unit for Planning (IMRUP), which actually started functioning on the very first day of the year under report, has already carried out investigations in regard to management, control and human relationship in three public enterprises, namely, the Mysore Iron & Steel Works, the Hindustan Shipyard (Private) Ltd., at Vishakapattam and at the Sindri Fertilizers & Chemicals (Private) Ltd., besides certain studies within the Insitute itself.

4. *National Sample Survey :* By far the most important project in the way of economic surveys, now under way in the country, is the National Sample Survey, a countrywide fact-finding agency operated jointly by the Government of India and the Institute. The surveys are organized in rounds— roughly two every year. During the year under review, the 10th round of the Survey, started in the preceding year as well as the field-work for the 11th round were completed. The 12th round was started early in 1957.

5. *National Sample Survey Review Committee :* A Committee, with Sir Ronald A. Fisher as Chairman, and Messrs. Morris Hansen (USA), Arthur Linder (Switzerland), Tosio Kitagawa (Japan) and Frank Yates (UK) as members, which had been set up with the approval of the Government of India, assembled in Calcutta in December 1957 and examined the work of the National Sample Survey and submitted its report in January 1957. The Committee appreciated the foresight shown in setting up this fact-finding agency and, while commending generally the conduct of the survey, made many valuable suggestions for its improvement and indicated lines of future development. The Institute in due course passed on this report with its own observations to the Government of India.

6. *Research and Training :* Besides carrying out theoretical research at a high level, the RTS has made it its duty to turn out trained personnel for statistical work. The number of persons who received training during the year under report was 350. Training in statistics is also imparted by the International Statistical Education Centre run jointly by the International Statistical Institute and the Indian Statistical Institute. Since its inception in October 1950, the Centre had conducted 10 terms of training upto April 1957 for 231 participants from 17 Asian countries.

7. *Surveys, Reports and Research Papers :* Ten sample surveys and pilot studies were conducted during the year, the total number till the end of the year being 185. The

number of reports on large-scale sample surveys submitted during the year was 9, bringing the total till the end of the year to 64. Altogether 48 working papers were drafted for circulation during the year under review, the total number of such papers circulated till the end of the year was 144.

8. The research papers and reports submitted by the workers during the year numbered 35, making a total of 577 of such papers and reports published till the end of the year.

9. The Research and Training School also brought out during the year 4 occasional publications.

10. *Visit to Japan and China:* A delegation consisting of Sri Pitambar Pant, Sri R. Natarajan and Sri M. Mukherjee visited China and Japan in the summer of 1956. Professor Mahalanobis was to have joined the delegation in Japan on his way back from the USA, but unfortunately he was suddenly taken ill in the USA and in his absence Sri Pitambar Pant acted as the leader of the delegation. Leaving Calcutta on 10 June 1956, the delegation spent about a month each in Japan and China and returned to Calcutta on 19 August. The delegation studied various aspects of planning in both the countries and brought back valuable materials on the subject.

11. Before passing on to the sectional reports mention should be made of two notable events of the year under review:

a) *Visit of the Chinese Premier:* Premier Chou En-lai, accompanied by Mr. Ho Lung, Vice-Premier of the People's Republic of China, visited the Institute on 9 December 1956 and was taken round the different departments in the activities of which they showed keen interest. Mr. Chou En-lai observed during the visit that technical cooperation between China and India in the field of statistics and planning would be very welcome. His visit was followed by that of a delegation from China led by Mr. Wang Szu Hua, Vice-Director of the Chinese Central Bureau of Statistics, who came to India as guests of the Institute and stayed in the country from 11 December 1956 to 15 January 1957 studying sampling and other statistical methods in use in India. It is expected that these visits will strengthen the ties between India and China.

b) *Conferment of Doctorate (Honoris causa):* In connexion with the Centenary Celebrations of the Calcutta University the degree of Doctorate of Science (*honoris causa*) was conferred among others on the following persons who are intimately connected with the Institute: Sri Chintaman D. Deshmukh, President; Professor Satyendranath Bose, Vice-President; Professor Prasanta Chandra Mahalanobis, Director and Professor T. Kitagawa, a Visiting Professor and a member of the Institute Council.

A brief history of the Institute is given in Appendix 1.

1. RESEARCH AND TRAINING SCHOOL

1. The main activities of the Research and Training School (RTS) fall under the three heads : (i) research ; (ii) consultation and (iii) training. A brief account of the work is given below. Full details will be found in a separate report of the Research and Training School.

(i) Research

2. The research programme of the RTS for the year covered both statistical theory and the logic of statistical inference as well as the development and use of statistical tools for application to specific problems. On the theoretical side, the studies included design of experiments, estimation, tests of hypotheses, stochastic processes and problems of distribution. On the applied side, the Psychometric Unit of the RTS was engaged in developing techniques for selection of personnel both within and outside the Institute, the preparation, scoring and validity of psychological tests and other studies. The Anthropometric Unit was busy with the standardization of measurements on the living. The Biometric Unit continued its work on certain bio-chemical studies, blood group tests and other problems. A summary of research work in theoretical statistics together with a full list of papers contributed, is given in Appendix 3.

3. The RTS also collaborated in the researches in other sections, such as, the design of sample surveys and other studies in the National Sample Survey (NSS) and associated projects; mathematical models for economic planning and other problems in the Planning Division; problems in high speed computation in the Electronic Computer Laboratory and industrial sampling and associated problems in the Statistical Quality Control Units of the Institute. A fuller account is given in the separate Report of the Research and Training School.

4. *Research Seminars :* Besides the seminars periodically conducted by the RTS for exchange of ideas among the staff and students, several courses of lectures on recent advances in mathematics and statistics were arranged during the year. Among them, mention may be made of a course on biometry by Professor J. B. S. Haldane of the University College, London;* on statistical inference by Professor G. A. Barnard of the Imperial College of Science and Technology, London; and on operational research by Dr. R. L. Ackoff of the Case Institute of Technology, USA.

(ii) Consultation

5. *Scientific Enquiries :* As in previous years, the RTS continued during the year to provide consultation service for all scientific workers in the country. Advice was given in response to various enquiries on the planning of investigations and analysis and interpretation of data from research workers in various fields, government departments and business firms. Further details are given in Appendix 4.

---

* Professor Haldane joined the ISI as Research Professor in July, 1957.

(iii) Training

6. An increased demand for training in statistics has led to an expansion of the scope of the training programme of the school, both in regard to the number of courses conducted and the number of trainees admitted to each course.

7. The short-term (6 months) Statistician's Course, which was introduced in 1955 for persons who are already in employment and who intend acquiring a knowledge of basic statistical methods for use in their own work, was extended to one year spread over two sessions of six months each (junior and senior courses), the senior course providing for specialization in one or two fields of application.

8. Another new feature was the introduction of short intensive courses. Two such courses were arranged during the year under review : one was a six-week course on the theory and practice of the design of experiments, for which 12 out of 84 applicants were selected; the other was a three-week course on operations research, which was attended by 30 technical workers drawn from different occupations.

9. The following courses of training (including the new courses mentioned above) were conducted during the year, the number of trainees being mentioned within brackets in each case :

> (*a*) Three-year Statistician's Course in collaboration with the Central Statistical Organization (CSO) of the Government of India (75); (*b*) Short-term Statistician's Course (junior and senior) (84); (*c*) Officers' Training Course (jointly with the CSO) (25); (*d*) Training of individual officers on deputation (6); (*e*) Computer's Training Course (junior and senior) (118); and (*f*) Short Intensive Courses on special subjects (42).

In addition, the staff of the RTS participated in the teaching work of the International Statistical Education Centre (ISEC).

10. *Advanced Studies :* Research scholarships were awarded to 13 students for carrying out work on subjects such as Advanced Probability, Stochastic Processes, Statistical Inference, Multivariate Analysis, Statistical Quality Control, Psychometry, Biometric Methods, etc. Three members of staff on the RTS were admitted to the degree of D.Phil. of the Calcutta University during the year; a fourth member submitted his thesis for the degree. Three research students submitted theses for the Associateship of the ISI. Details are given in Appendix 5.

2. INTERNATIONAL STATISTICAL EDUCATION CENTRE

1. The International Statistical Education Centre (ISEC), Calcutta, was opened in October, 1950 under the auspices of the UNESCO. It is being run jointly by the International Statistical Institute and the Indian Statistical Institute with the support of the UNESCO and the Government of India. The Centre provides courses of training in theoretical and applied statistics at various levels to selected participants from countries of the Middle, South and Far East.

2. During the year under review, important decisions were taken by the Government of India regarding the future of the ISEC. During the past six years of its existence,

the work of the centre was reviewed every year and grants for its maintenance made by the Government of India on a year-to-year basis. This has made it difficult to plan the programme of the centre in advance. The Government of India have now kindly sanctioned grants to the ISEC upto the end of the Second Five Year Plan period (1960-61).

3. Since its inception upto April 1957, the centre has conducted 10 terms of training of either six months or nine months for 231 participants from the following 17 Asian countries : Afghanistan, Burma, Cambodia, Ceylon, India, Indonesia Iran, Iraq, Japan Malaya, Nepal, Pakistan, Philippines, Singapore, Syria, Thailand and Vietnam.

4. The 9th term terminated about the middle of April 1956. The 10th term, which started on 16 July 1956, terminated in April 1957. During this term, 30 candidates were selected from 38 applicants and 28 candidates from 8 countries actually joined and received training; 18 students came from outside India: 2 each from Burma, Ceylon, Philippines, Singapore and Thailand and 4 each from Japan and Pakistan. A full list of students is given in Appendix 6.

5. *Instruction :* The course of training during the 10th term consisted of lectures, laboratory work, seminar discussions and on-the-job training. The first four months were devoted to theoretical work involving about 175 lectures and 400 hours of laboratory work. About a month was spent in imparting training in statistical projects in the ISI. Another 2 months were set apart for a specialization course, for which students were divided into appropriate groups. This was followed by a course of training in official statistics for a month in the CSO at New Delhi. Instruction at the Centre was provided by the staff of the RTS and project divisions of the ISI, statisticians and economists in the CSO and from various ministries and departments of the Government of India and visiting teachers from abroad.

6. In addition to their regular work, the students of the ISEC had an opportunity to attend various lectures and seminars held in the ISI. For example, they attended the series of seminars in the ISI which formed a part of the 25th Anniversary Celebrations of the Institute.

7. *Fellowships :* The Government of India have awarded, under the Technical Cooperation Scheme (Colombo Plan), 102 fellowships to ISEC trainees during the first nine terms and to 17 students during the 10th term.

8. *Certificates :* Out of the 29 students who joined the 10th term, one from India could not complete the course. The remaining 28, having satisfactorily completed the course, were awarded certificates at a function held on 5 April 1957.

9. *Students' activities :* The students were taken out on an educational tour to places of historical interest, such as, Agra, Banaras, Gaya and a Community Project Centre at Nilokheri near Delhi. Excursions were also organized to places of interest in and around Calcutta. A souvenir volume covering the 10th term has already been prepared and is available for circulation.

10. *Symposium on the teaching programme of the ISEC :* Advantage was taken of the presence of statisticians from various countries during the 25th Anniversary Celebrations of the ISI to organize a symposium on the teaching programme of the ISEC, which

was held on 26 December 1956. Among those who participated were Mr. A. Afzal (CSO, Pakistan), Mr. S. K. Banerjee (ISEC Student, Calcutta), Dr. C. Chandrasekharan (Ministry of Health, New Delhi), Mr. Q. M. Hussain (Dacca University), Prince M. C. Athiporn P. Ksemasri (CSO, Thailand), Mr. Faiz El. Khuri (ISEC, Beirut), Dr. N. Keyfitz (Bureau of Technical Cooperation, Colombo), Mr. E. Lunenberg (International Statistical Institute, The Hague), Dr. A. Matthai (Indian Statistical Institute), Mr. R. Minobe (Bureau of Standards, Tokyo), Dr. P. B. Patnaik (CSO, New Delhi), Dr. B. Ramamurti (CSO, New Delhi), Dr. C. R. Rao (Indian Statistical Institute), Mr. S. Sen Gupta (Eastern Railway, Calcutta), Dr. M. Ziauddin (Punjab University, Lahore) and Professor A. Linder (University of Zurich). Professor Linder was the Chairman of the Symposium. On the basis of discussions at the Symposium, a committee on teaching courses was appointed consisting of six members nominated from among the participants. The committee made recommendations on different types of courses, curricula of the courses, system of examinations, award of certificates and the selection procedure of candidates. Some of the recommendations will be given effect to in the 11th term.

3. PROFESSIONAL EXAMINATIONS

1. Since 1938 the Institute has been conducting professional examinations on an all-India basis for the award of certificates for computers and field investigators and of post-graduate diplomas for statisticians. These examinations are conducted by an Examinations Committee consisting of statisticians and economists selected from all over India. The examinations are open to all candidates who satisfy certain educational and/or training requirements. Paper-setting and examination of scripts are done by examiners selected by the Examinations Committee from both inside and outside the Institute. These examinations are of a professional nature.

2. Apart from the examinations mentioned, the Institute also awards its Associateship and Associate Fellowship to candidates on the basis of certain regulations which are somewhat similar to those recently initiated by the Royal Statistical Society and are conducted by a standing committee appointed under the regulations.

3. For statistical and computational posts in some Government and private institutions, the Institute's diplomas and certificates are given as much weight as university degrees.

4. During the year under review, professional examinations were held twice, in September 1956 and in March 1957. In the September examinations, there were 121 candidates for Statistician's Diploma, 148 candidates for Computer's Certificate and 72 candidates for the Field Survey Certificate. The examinations were held at 8 Centres: Bangalore, Bombay, Calcutta, Giridih, Lucknow, Madras, New Delhi and Poona. 50 candidates were successful in at least one paper of the Statistician's Diploma Examination; the corresponding figures for Computer's Certificate and Field Survey Certificate were 136 and 24 respectively. The March examinations commenced during the last week of the month; the number of candidates registered were 133 for Statistician's Diploma, 180 for Computer's Certificate and 62 for Field Survey Certificate. The results of the examinations published during the year are given in Appendix 7.

4. NATIONAL SAMPLE SURVEY

1. The National Sample Survey (NSS) is a country-wide fact-finding agency operated jointly by the Government of India and the Indian Statistical Institute (ISI). The Field Branch is under the control of the Cabinet Secretariat and the statistical work is done in the Institute. Although a large part of the work done in the Institute is of an operational nature, a good deal of research and experimental studies are being carried on in the Institute to improve the design of the survey and the techniques of collection, processing and analysis of the data. Only a general account of the work done during the year under review is given here. More detailed information will be found in Appendices 8, 9, 10 and 11.

2. The survey is organized in rounds, roughly two every year. The 10th round of the survey, which commenced in December 1955, was completed in May of the year under review. Immediately after this, a special survey was carried out at the request of the Ministry of Food and Agriculture.

3. The field work for the 11th round was started in August 1956 and completed in February 1957. In addition to standard items, detailed information was collected, at the desire of the Ministry of Labour, regarding agricultural labour, such as size and composition of agricultural labour households, employment, wages, income, expenditure, etc. Some information was also collected in this round, at the instance of the Standing Metric Committee, on different systems of weights and measures actually used in the country.

4. The 12th round of the survey, which started early in 1957, was scheduled to be completed by August 1957. There was no substantial change in its scope apart from introduction of a new schedule designed to ascertain the output of milk in the country. The agricultural labour enquiry was continued in this round and certain improvements were incorporated in the schedules on the basis of discussions with the Agricultural Labour Unit of the Labour Ministry. The field training conference for this round was held in February 1957.

5. The 5th round of the sample survey of manufacturing industries (SSMI), covering the reference year 1955, started during the year under review. A survey of some scheduled industries, listed in the First Schedule of the Industries (Development and Regulation Act, 1951), was conducted for the first time along with the 4th round of the SSMI. The scope of the SSMI was extended in the 5th round to cover nearly all the industries under the above Act. Several new items were introduced in the schedule with a view to collecting information regarding various labour welfare measures. Information was also collected on the occupational structure of employment, with particular reference to skilled personnel, which was urgently required by the Planning Commission.

6. Considerable progress was made in the preparation of reports. Ten draft reports were got ready, out of which nine were submitted to the Government for obtaining clearance for final release; clearance was received for the publication of four reports and four reports were in the press. Apart from these reports, information collected through the NSS was supplied to a number of Government and other agencies.

7. A large part of the processing work of the NSS is done mechanically in adequately equipped machine tabulation sections located at Baranagar (Calcutta), Giridih and Delhi (in the Army Statistical Office). With the addition of one new machine unit each at the

3 stations, the present equipment comprises 2 IBM electronic statistical machines, 23 sorters, 15 tabulators, 5 multipliers, 5 collators, 21 reproducers and gang summary punches of the Hollerith (IBM and BTM types) and 2 sorters, 1 tabulator, 1 interpolator and 1 multiplier of the Powers-Samas type.

8. Two tabulators (type T. 5), 4 sorters, 20 automatic electronic key punch duplicators and 6 electrical verifiers, received from the USSR through the UNTAA, were duly installed. These new machines were serviced and maintained by the Institute itself. The machines, which were rented from the Hollerith, IBM and Powers-Samas organizations, however, continued to be serviced by the respective firms.

9. Special training courses were arranged with a view to an expansion of the operating staff. Out of 25 candidates, 12 successful candidates were absorbed at Baranagar, 5 other candidates were recruited, trained and absorbed at Giridih and 18 other trainees were selected at Giridih for training. During the year, 7 trained machine operators left the Institute, which made it necessary to organize a second training course to fill the vacancies.

10. In addition to the operational work, researches and investigations were taken up in the Research Unit and the Special Unit. The Research Unit was mainly engaged in studies for the improvement of the design of successive rounds of the NSS. This Unit made 22 pilot studies (based mostly on small self-weighting samples) to assess the quality of primary material and the effectiveness of methods of collection of data and to explore profitable lines of tabulation and analysis. The small size of the pilot samples and the simple method of estimation used made it possible for the technical workers to examine the material critically, judge its quality and recommend methods for improvement.

11. The work in the Special Unit related mainly to land utilization and crop surveys. Among the investigations taken up during the year mention may be made of experiments on land surveys in areas (near Giridih) carried out in August–October 1956, a study of the distribution of harvestable fields on different days of a crop season in three areas in West Bengal (August–October 1956), experiments with sample-cuts of various sizes and shapes for estimating the yield of crops per acre; comparison of different methods of collecting information on crops (a team visiting 6 sample-villages per month on foot, a team visiting 12 sample-villages per month on bicycles; and a team visiting 24 sample-villages per month in a motor car), and a study to find out whether the rate of yield per acre of crops varies with the distance of the plot from the nearest road.

12. Several other studies on socio-economic conditions and on survey techniques were carried out during the year, such as a re-survey of 410 households in Calcutta (originally surveyed in April–September 1952), to study the change in the pattern of consumption, in which only 260 of the original households could be traced; a study to find out how information on consumer expenditure varies when the data are collected from different respondents; an experimental study of the level and character of trading activity in West Bengal; a pilot survey with pre-coded schedules; a complete enumeration of population on census slips in 3 census blocks by one investigator in November 1956, which was repeated later by a second investigator to study the accuracy of complete enumeration; a study of biases in enquiries on household consumption in Calcutta which was started in January 1957.

13. A number of agricultural field trials were also conducted in the Institute's own farm at Giridih to study the effect of different numbers of seedlings per hole, different

spacings between holes and different treatments of fertilizers in the case of paddy; the effect of fertilizers on sugarcane; and the progressive maturity of *aman* paddy transplanted according to the Japanese method.

14. The Demographic Unit conducted a preliminary survey of family planning in 49 census blocks of Calcutta in the course of which married persons in 8000 households were interviewed. A somewhat similar survey was made of the extent to which the benefits of hospital services were actually availed of in Calcutta by interviewing a sample of ex-patients of hospitals.

15. *Review Committee :* Mention has been already made of the NSS Review Committee in an earlier section. The NSS staff derived much benefit on the technical side through personal discussions with distinguished members of the Committee. The recommendations made by the NSS Review Committee were examined by the NSS staff and the views of the Institute were submitted to Government.*

5. ELECTRONIC COMPUTATION

1. An important part of the research activity of the ISI in recent years has been the development and utilization of digital electronic computation. A general purpose electronic computer, HEC-2M, which was received in February 1956, started work under full load during the year under review. A fair number of computational problems submitted by the departments within the Institute and by other scientific institutions and Government departments were solved on this machine. Mention may be made of the following institutions from whom problems were received :

> (1) The Andhra University, Waltair; (2) The Indian Institute of Technology, Kharagpur; (3) The Indian Institute of Science, Bangalore; (4) The Indian Association for the Cultivation of Science, Calcutta; (5) The Physical Research Laboratory, Ahmedabad and (6) The Tata Institute of Fundamental Research, Bombay.

2. The Institute staff not only maintained this complex and delicate machine but also spotted out the most frequent causes of breakdown and made a large-scale modification of the machine with a view to eliminating these causes. Some useful monitoring devices have already been added to this machine. Experiments have also been started to replace the mechanical switching device of HEC-2M machine by an all electronic switch. Some experiments on magnetic gates have also been made.

3. A large number of electronic instruments, consisting of 10 synchroscopes, 6 Q-meters, 6 pulse generators, 10 vacuum tube volt-meters, 4 quartz calibrators, 8 insulation testers, 10 ampere-volt meters, 8 universal resistance bridges, 6 tube testers, 15 volt-ohm meters, 10 signal generators and a few audio-signal generators, were received during the year from the USSR under the United Nations Technical Aid Programme. The Russian machines received were maintained by the staff of the section and regular training on the plug-board setting and operations of these machines was also undertaken. A number of simple computational problems were solved on the Russian Tabulator by an ingenious setting of the plug-board, thus relieving the electronic computer of some of its load.

---

* Since then the Institute has set up a Special Committee consisting of Professor D. R. Gadgil, Sri J.J. Anjaria and Sri N. C. Chakravarti.

4. Transferring data from 65-column Powers-Samas to 80-column IBM cards presented a serious problem in the Institute. (A machine for the reverse operation, that is, for transferring information from IBM cards to Powers-Samas cards, is available in the market). The design of a scrap reproducer was modified to solve this problem and a few thousands of cards were reproduced successfully in the new machine.

5. A new method of finding the reciprocal of a matrix of a large order on electronic computers was evolved during the year. If the matrix happens to be singular, the method can dictate the presence of singularity and the coefficients of the linear dependence of the column (or row) vectors can be determined directly from the results of previous computation. The method has been successfully applied to the problem of inverting a matrix of order 34 which arose in connexion with an econometric problem. Several research papers on numerical analysis, applied statistics and electronic engineering were completed by the members of the staff.

6. Mr. Konoplev, a senior engineer from the USSR deputed by the UNTAA and Academician Dorodnitzin of the USSR Academy of Science, gave lectures during the year on the electrical system of the Russian tabulators and on the organization and activities of the Computing Centre of the Academy respectively. Mr. S. N. Alexander, Chief of the Data Processing Systems Division of the National Bureau of Standards, USA, whose visit was sponsored by the American Government, delivered 4 lectures on the research and development of high speed electronic computers in the National Bureau of Standards and also participated in numerous technical discussions with the members of the Institute.

7. Four engineers, deputed by the Departments of Radiophysics and Electronics of the Calcutta University, successfully completed training in electronic computation in the Institute during the year.

6. STATISTICAL QUALITY CONTROL

1. The Institute has been advocating since 1935 the use of Statistical Quality Control (SQC) in the country and since 1950 has been maintaining three SQC Units in Calcutta, Bombay and Bangalore with the help of grants from the Government of India. The programme included (a) introduction and maintenance of SQC in factories, (b) training and (c) promotion of SQC in general. Service, which is partly subsidized, is given to member firms on a fee basis. At the end of March 1957, there were 44 member firms on the roll as against 32 in March 1956.

2. *SQC Policy Advisory Committee :* It would be recalled that the SQC Policy Advisory Committee was set up in 1954 under the Chairmanship of Sri C. D. Deshmukh, then Finance Minister of the Union Government. In the year under review at a meeting held on 7 January 1957 Sri Deshmukh suggested that as he was no longer connected with the Government and as the work of the Institute had also been transferred from the Ministry of Finance to the Cabinet Secretariat, the Committee might be reconstituted under appropriate Government sponsorship. All members present agreed to his suggestion and the Committee has since been reconstituted under the sponsorship of the Cabinet Secretariat with the following membership : Sri C. D. Deshmukh, Chairman; Sri K. C. Neogy, Prof. P. C. Mahalanobis, Sri G. D. Somani, M.P., Sir Shri Ram, Sri M. G. Kotibhaskar, Sri Vikram Sarabhai, Sri S. C. Jain, Dr. D. S. Kothari, Sri S. Ananta

Ramakrishnan, Dr. Lal C. Verman, Dr. M. S. Thacker, Sri S. S. Khera, Sri Vishnu Sahay, Sri P. A. Gopalkrishnan, Dr. A. Nagarajarao, Members and Sri Pitambar Pant, Member-Secretary.

3. Mention was made in the previous year's report of the decis ion of the Ministry of Defence to set up a Statistical Quality Control Unit for the Ordnance Factories which would operate under the guidance in policy matters of the SQC Policy Advisory Committee. This Unit, it is satisfactory to note, achieved significant results in the very first year of its existence.

4. *SQC Unit, Calcutta :* The Unit had 10 member factories on its roll as against 8 in the previous year. Pilot investigations were conducted in 3 other factories and introductory visits were paid to 4 others. The first series of courses for local industry was organized during the year which stimulated a demand for SQC service. An advanced Seminar, held in September-October 1956, was addressed by Professor Barnard of the Imperial College of Science and Technology, London. The introductory course on SQC attracted 23 trainees from 16 factories while 16 trainees from 10 factories attended the advanced course.

5. Arrangements were made during the year for close collaboration with the Research and Training School (RTS). Three RTS trainees received practical training in SQC with the help of the Unit and the staff of the Unit participated in the RTS course on design of experiments. Lectures, talks and discussions were also organized as a part of the promotional activities of the Unit.

6. Fresh recruitments of graduate assistants and technical assistants were made in the Unit during the year; and were placed in factories under senior technicians. Training courses were organized for the benefit of managerial and technical staff in the industry. Special lectures by experts were also arranged from time to time. All staff members were encouraged to attend a course of 3 lectures on Operations Research by Mr. R. L. Ackoff already mentioned.

7. Mr. D. J. Desmond, who joined the Unit in December 1955, was away for seven months in 1956. The Unit had the benefit of guidance from Mr. A. L. Raich during the winter of 1956-57. Some experts who came by arrangements with UNTAA also helped the work of the unit during this period. The strength of the Unit was 10 technicians and 2 general assistants at the end of the year under review.

8. *SQC Unit, Bombay :* This Unit had 16 member factories on its roll as against 14 in the previous year. Apart from servicing the member factories, pilot studies were conducted in the Bombay Telephone Workshops and some testing work was undertaken on the *Ambar Charkha* yarn at the request of the All India Khadi and Village Industries Board.

9. Lectures, talks, discussions and an exh ibition of "SQC in Industry" were organized by the Unit by way of promotional activities. Professor Barnard addressed a meeting of statisticians representing different industries on 27 October 1956. Mr. A. Zelinovsky, Deputy Chairman of the USSR Planning Commission and Mr. A. Polosukhine, Industrial Specialist to the USSR Economic Commission, also addressed a meeting held in their honour. Sri C. D. Deshmukh, President of the Institute and Chairman, Policy Advisory Committee on SQC, met a group of industrialists on 22 February 1957. Promotional visits were also paid to 5 factories which subsequently enrolled themselves as members.

10. Three apprentices were trained in the Unit during the year under review. Five statisticians who joined member factories during the period under review were also trained in the Unit. The staff consisted of 11 technicians, 2 persons for general administration and 4 persons for miscellaneous and other work.

11. *SQC Unit, Bangalore :* The Unit continued to serve 15 member establishments in Bangalore, Coimbatore, Madras, Salem and some other places. Promotional activities were undertaken in and around Bangalore. A number of talks, discussions and meetings, covering theoretical and practical aspects of SQC, were organized for the benefit of technicians, managers and other executives of several establishments. Mr. A. L. Raich gave a talk on "Cost Function in Quality Control" on 16 February 1957 and Mr. D. J. Desmond a talk on "Economics of Sampling" on 28 August 1956. Professor Srinagabhushana of the Unit contributed two technical papers on "Quality Control in Textile Industry" and "Testing of Textile Materials and Statistical Approach to Testing of Textile Materials."

12. Three assistants were recruited and given training during the year. The Unit also gave training in SQC to engineers belonging to the Industrial and Management Research Unit for Planning of the ISI by conducting a short term course in March 1957. A 10-day course on SQC was conducted by the Unit in Madras in summer, 1956 for training factory technicians. A discussion meeting was held in Madras in March 1956 in which managing directors and managers of 26 firms participated. At their request, visits were arranged for some of the establishments to undertake pilot studies to explore the possibilities of applying the methods of SQC.

13. The strength of the technical and administrative staff of the Unit were respectively 11 and 5. Mr. Desmond worked for the Unit during the period April to August 1956.

14. *SQC Sub-unit, Coimbatore :* This Unit had 10 factories on its roll as against 5 last year. Besides servicing the member factories a number of talks and discussions were held on theoretical and practical aspects of SQC.

7. WORK ON PLANNING FOR NATIONAL DEVELOPMENT

7.1. Economic Planning

1. It will be recalled that on 3 November 1954, Prime Minister Jawaharlal Nehru inaugurated studies relating to planning for national development in the Institute. The "Draft Plan Frame", which was submitted in March 1955, was accepted as a basis for the formulation of the Second Five Year Plan. Professor N. Kaldor's work on taxation, which had been undertaken on behalf of the ISI, received serious attention in Government circles and was drawn upon in framing the budget of the Central Government. A small Operational Research Unit (ORU) was established and an Economic Wing(EW) was started in 1955-56. Work on planning in the Institute is being done from the very beginning in collaboration with the Planning Commission, the Central Statistical Organization and the Economic Division of the Ministry of Finance. Attempts were made in various directions to organize the work with some emphasis on long-range or perspective planning.

2. Some studies were taken up on certain aspects of long-term planning and an attempt was made to prepare an annual plan and, subsequently, a projection covering

fifteen years yielding output targets for 1970-71. These results were not intended to have any bearing on immediate policy decisions but were exercises aimed at developing appropriate planning techniques. Another study on long-term planning in more operational terms was under way towards the end of the year. Some studies were also undertaken on problems in the transportation sector.

3. A section was engaged in studies on inter-industry relations. A (34×34) table for 1951-52 was converted into physical terms and the coefficients of the transformed table were used to set up a method for estimating prices warranted by existing institutional factors. A much larger table (116×116) for 1953-54 was completed and a similar table for 1954-55 was under construction. Meanwhile, attempts were made to present the input coefficients in the 1953-54 matrix with their errors of estimation. Preliminary studies were made on the estimation of levels of activity and levels of final demand. An experimental model for estimating the time-path of levels of activity in response to the assigned time-path of final demand was set up. A new method was developed for the solution of linear programming problems; as no rigorous proof could be worked out, studies were conducted to establish its validity.

4. A third section worked on miscellaneous economic problems. A preliminary study was completed on the degree on monetization in final transactions at the instance of Dr. Wilfred Malenbaum of the Massachusetts Institute of Technology. Other studies were concerned with labour productivity, efficiency in relation to size of establishments, direction and pattern of foreign trade, availability of investible surplus excise taxes, regional variations of prices and the consumption pattern of the workers of the Institute. A team of workers cooperated with the NSS to prepare schemes of intensive surveys relating to employment, consumption, industrial output and land utilization; they also helped in the analysis and interpretation of NSS data.

5. Working papers on various subjects were prepared during the year under review; a list is given in Appendix 12.

6. A group of the workers in ORU and EW were posted in Delhi and worked in close collaboration with Government departments. Studies were undertaken on several aspects of technical manpower and perspective planning by a group associated with the Planning Commission and the CSO. Work was also undertaken on a form for reporting projects included under the plan in collaboration with the CSO.

7. A notable feature of the work on planning in the ISI was lectures by and discussions with foreign scientists whose views and criticisms were of great value to the workers of the Planning Division. Special mention may be made in this connexion of Professor Charles Bettelheim (France), who spent nearly one year in the Institute with an assignment from UNTAA; Sir Robert Hall (UK); Professor Simon Kuznets (USA); Mr. Alvin Mayne (Puerto Rico); Mr. A. Zelinovsky and Mr. A. Polosukhine (USSR) and Mr. J. P. Miller (USA).

8. A delegation from the ISI visited China and Japan during summer 1956 to study planning methods among other things. Apart from taking copious notes on the basis of discussions, the delegation managed to collect a good deal of valuable material in Chinese and Japanese. For translation of material of interest to planning, the scope of the Translation Unit set up under the Library was extended to cover Chinese and Japanese languages.

7.2. Kalyanasri

1. A special unit called *Kalyanasri* was started on 28 February 1956 to study on an experimental basis (i) the extent to which production in household enterprizes and crafts could be increased, if the supply of raw materials and marketing of goods could be assured; (ii) the time required to train a person to enable him to earn a reasonable remuneration and (iii) in what way and to what extent similar units can be established under usual economic and social conditions in Indian villages or urban areas.

2. The headquarters and urban centre are located in some cottage built on the grounds of the Institute with equipment and some training staff provided by the Institute. The first batch of trainees consisted of about half a dozen girls, mostly displaced persons from East Pakistan who were without work, but the number rose rapidly to about 50 by August 1956 when the programme was more or less stabilized. Some of these workers had training in cottage industries and crafts and were employed on a piece-rate basis. Others had to be given some training. The Institute provided a small staff to carry out statistical and economic studies relating to the activities of the Unit.

3. In addition, three villages were selected at a distance of about 15 to 20 miles from Calcutta, two in 24-Parganas and one in Howrah districts. A preliminary survey of all households engaged in cottage industry or crafts was carried out in each of these villages with a similar survey of a sample of other households which were selected at random. From the industrial households 15 were selected to carry out their respective crafts and enterprizes under experimental conditions with some assistance from the Institute.

4. Two other rural centres were established in Midnapore district to carry out similar studies in among certain backward people under the guidance of Sri Nirmal Kumar Bose of the Calcutta University.

5. Two types of data are being collected. One type relates to outlay on implements, input of raw materials, labour, and other items of cost of production, sales, etc., to study the requirements of capital in relation to gross and net value of output, volume of employment and also the effects needed for organization, training and supervision. The other type of data relates to the level of living of the households under observation and the effect of the additional earnings on the pattern of consumption to find out what new demands were generated by the new income.

6. In March 1957 there were 50 workers on the roll of the urban centre at the headquarters of *Kalyanasri*, of whom 48 were women; 9 were engaged in spinning on Ambar Charka, 13 in weaving with semi-automatic flyshuttle and pit looms, 6 in tailoring and 8 in weaving Manipuri-style of fabrics and other artistic crafts. All the workers were paid on a piece-rate basis, excepting 7 who were engaged in training and supervision. In the rural centres the development programme has been taken in hand since the close of the year under review except in Midnapore where a certain amount of work was done during the last part of the year.

7.3. Industrial and Management Research Unit for Planning (IMRUP)

1. The Council of the Indian Statistical Institute approved in March 1956 a scheme for the setting up of an Industrial Management Research Unit for Planning (IMRUP) with headquarters at Bangalore to study the problems of industrial management in both public

and private sectors. It started functioning from 1 April 1956 with two Directors (Dr Bhola D. Panth and Sri D. P. Basu) and a skeleton clerical staff. The third Director (Sri R. Natarajan) joined in May 1956.

2. The work of the IMRUP began with a preliminary symposium on the "Organization and management of public enterprizes" held in Bangalore in which representatives of industry and Government of India participated.

3. Investigations were carried out in the fields of management, control and human relationship in respect of the Mysore Iron and Steel Works at Bhadravathi and a report on the basis of this study was submitted. Similar studies were in progress in the Hindusthan Shipyard (Private) Ltd., Vishkapattam and Sindri Fertilizers and Chemicals (Private) Ltd., Sindri. A report was also prepared on the work of the ISI itself.

4. The work programme of the IMRUP was reviewed during the the middle of the financial year and an expansion of the scale of operation was decided upon. Eleven engineers were recruited towards the end of the year and a training programme was arranged for them, which included attendance at a course on "Operations Research" conducted by Dr. R. L. Ackoff from 28 January to 9 February 1957 at Bangalore.

5. Sri R. Natarajan visited Japan and China (June-August 1956) as a member of the Institute's delegation and prepared a report on "Industrial Planning and Growth in China."

6. IMRUP is yet in its formative phase. Economic development involves rapid industrialization in the country, particularly in the Government sector and the success of the plan will largely depend on the efficient management of the industrial concerns. IMRUP is expected to study and devise efficient methods of management in individual cases and offer diagnostic and consultation service when there is need.

7. The headquarters of IMRUP are situated at Bangalore, although the Directors frequently visit the other offices of the Institute as well as industrial enterprizes. The staff at the end of the year under review consisted of 3 Directors and 26 other members including 11 engineers.

7.4. Geographical Projects

1. Located at IMRUP headquarters are also the offices of Dr. A. T. A. Learmonth, who came from the Department of Geography, Liverpool, to work with the Institute on 6 August 1956 with an assignment from the United Kingdom under the International Technical Assistance Scheme of the Colombo Plan. He is engaged on a pilot project for Regional Planning in which an attempt is being made to combine the geographical with the statistical methods, thus leading to a new technique for economic surveys. Dr. Learmonth's immediate objective is a micro-survey of the State of Mysore for the production of a series of maps to indicate the physical as well as the economic features. Dr. Learmonth is also cooperating with a macro-survey of the surrounding regions in Andhra, Mysore and Kerala, so that the new method may be tried on both macro and micro scales. A committee consisting of officials of the Planning Commission, CSO and IMRUP assists Dr. Learmonth in his work.

2. Before initiation of the above project Dr. O. H. K. Spate, a well-known geographer, was in India for about a month and a half in October-November 1956, as a guest of the Institute. A three-day conference was organized in the Planning Commission on 19-21 October 1956 in New Delhi. Sri Guljarilal Nanda, Minister of Planning, inaugurated the conference on 19 October 1956 and Dr. J. C. Ghose, Member, Planning Commission, Prof. P. C. Mahalanobis and officials of the Planning Commission participated in the discussion. Besides Dr. Spate and Dr. Learmonth, Sri Mani Mohan Mukherjee, Central Statistical Organization, New Delhi, Dr. B. Ramamurti, Central Statistical Organization, New Delhi, Dr. V. Nath, Director, Programme Evaluation Organization, New Delhi, Dr. V. L. S. Prakasa Rao, Senior Lecturer in Geography, University of Madras, Dr. G. K. Ghori, Professor of Geography, Maharaja's College, Mysore, Dr. O. P. Bharadwaj, Lecturer in Geography, Punjab University, Dr. R. L. Singh, Professor of Geography, Banaras Hindu University and Sri V. A. Phillippes, Delhi Town Planning Organization took an active part in preparing a broad scheme for the project. Subsequently another conference was arranged in the Institute in Calcutta on 24-25 November 1956 which was presided over by Prof. O. H. K. Spate and was attended by Dr. E. Ahmed, Ranchi College, Ranchi, Dr. S. M. Ali, Aligarh University, Prof. S. P. Chatterjee, Calcutta University, Prof. C. D. Deshpande, M. N. College, Visnagar, Prof. G. Kuriyan, Madras University, Dr. A. T. A. Learmonth, Liverpool University, U. K., Prof. P. C. Mahalanobis, ISI, Mr. M. P. Thakore, Punjab University College and Dr. R. L. Singh, Banaras Hindu University.

8. SERVICE UNITS

8.1. Machine Tabulation Unit

The activities of the Machine Tabulation Unit have been briefly mentioned earlier in this report under NSS (paragraphs 4.7 to 4.9) and Electronic Computer Section (paragraph 5.4). The strength of tabulation equipment at Baranagar, Giridih and Delhi, as in April, August and November 1956 are given in Appendix 13.

8.2. Library

1. The Central Library, which had so far been partially housed at 204 Barrackpore Trunk Road, was shifted to the main premises at 203 Barrackpore Trunk Road. It contained about 60,000 books and monographs towards the end of the year under review. Service branches were maintained as before in the City Office at 9B Esplanade East, Calcutta and at Giridih.

2. The Library acquired during the year 2873 volumes of books and monographs and received 1297 periodicals and annuals against 2717 volumes of books and monographs and 1190 periodicals and annuals last year. The total number of books, journals and other materials issued was 46,271 (to 995 persons) against 38,055 (to 758 persons) in the previous year.

3. Apart from routine work, other services rendered by the Library included : (1) bibliographical services; (2) a workers' circulating library; (3) news clippings; (4) translation from foreign languages and a language class in Russian; and (5) maintenance of records. Full details are given in Appendix 14.

8.3. Special Technical Unit

1. The Special Technical Unit (STU) was formed at the instance of the Director on 1 April 1955 with a small staff. Its object was to take up urgent work whenever necessary and to train up promising workers.

2. The demand increased rapidly and the STU had to be expanded considerably. An auxiliary Computing and Punching Unit, which had been established earlier, was first brought under the control of the STU and was completely integrated with the STU in 1956. During the last two years, about 50 workers were deputed or transferred after training in the STU to other departments. The STU prepared a draft report on the NSS data on investigators' time records (8th round) and some important tables with notes on the data relating to land utilization covering the 5th, 6th and 7th rounds of the NSS were under preparation at the end of the year. The STU was also responsible for the collection of data and various studies on cottage industries in the Kalyanasri.

3. Evening classes were organized to give systematic training to the STU workers in computation and project work.

8.4. Workshop

1. The Institute workshop functioned on an expanded scale during the year under review. The development section was responsible for the repair, maintenance and development work of the Electronic Computer Section and served other research needs. The Design and Drawing Office attached to the workshop was reorganized in April 1956 and acquired, during the period under review, complete drawing accessories including one ferro-printing machine made in the Institute.

2. A prototype desk calculating machine was manufactured and work in connexion with a second machine, with major alterations in the control arrangements, was also completed during the year under review.

3. The precision work for transferring information from 65-column Powers-Samas punched cards to 80-column IBM cards was successfully completed. The development work on the punched card sorter machine was well under way. The principles of other statistical machines were also being studied for developmental purposes.

4. During the year, 434 desk calculators were repaired and 6 desk calculators were reconditioned almost from scrap. The workshop also looked after the repair and maintenance of all motor vehicles of the Institute.

5. Maintenance and installation of electrical equipment developed rapidly ; to increase the capacity of the 200 k.v.a. transformer, two capacitators were installed. The Institute also acquired electric arc-welding plants, motors of various types, rectifiers, motor converters, portable electric tools, etc.

6. Air-conditioning being essential for the smooth operation of electronic and other precision calculating machines and also for the testing and maintenance of high precision machine parts, the installation and maintenance of air-conditioning and refrigeration units were taken up during the year. A maintenance shop with 3 workers was started for this purpose and installation of 4 machines with a capacity of 16 tons was completed.

7. During the period under review, the Institute received 218 pieces of Russian machinery through UNTAA valued at about 12 lakhs of roubles or about Rs. 14.4 lakhs approximately, bringing the total to date to 236 items valued at about 13 lakhs of roubles or about Rs. 15.5 lakhs. A large number of these units were installed during the year.

9. SOCIAL AND WELFARE SECTION

9.1. Medical Welfare Unit

1. In Calcutta and Giridih the benefits of the Medical Welfare Unit cover all workers of the Institute as also members of their families and include free consultation and service at the Unit, house visits to workers' residences at a subsidized rate, and prescriptions served on a cost basis. Arrangements have been made by the Institute for providing medical aid to Institute workers and guests in the New Delhi Office also, though no separate unit has been set up there.

2. The Calcutta Unit treated 10,803 patients against 8,819 in 1955-56; and supplied medicine worth about Rs. 13,500 to the workers against about Rs. 10,000 last year; gave six-monthly inoculations against typhoid and cholera and yearly vaccination against small-pox; the benefit of anti-malarial measures was extended to the entire field staff and the Unit also looked after the sanitary condition of the Institute campus, including hostels and residential quarters.

3. The Giridih unit treated about 2,000 patients and supplied medicine worth about Rs. 2,600.

4. There is also a Health Home at Giridih for convalescent workers, which was built on a plot of land received as a gift from Sm. Nirmal Kumari Mahalanobis partly with the help of a donation of Rs. 5,150 from the Soviet scientific delegation in 1955 and was opened by Dr. Frank Yates in March 1956.

9.2. Rehabilitation Loan to Refugee Workers of the Institute

A loan of Rs. 6,451 was granted to 10 workers of the Institute for house-building during the period under review. A total of Rs. 13,435 was spent for educational help to Institute workers during the year, the beneficiaries numbering 425.

9.3. Night School and Adult Literacy Drive

The Institute provided accommodation and some essential equipment to the Night School which made steady progress during the year. Out of the five students sent up for the School Final Examination, three proved successful. At the beginning of the new session, there were 106 students on the roll of the School, one-third of the students being workers of the Institute. Adult literacy classes were also conducted during the year with some success; the maximum attendance in the Hindi classes was 58.

9.4. Public Relations Work

Public interest in the Institute has increased considerably in recent years, particularly since work was started on national planning. Over 600 visitors were shown round the Institute during the year under review, and the list of visitors contained many eminent names. A Bengali translation of Professor Mahalanobis's radio talk on "Approach to Planning" was published during the year under review.

9.5. Guest House

The Institute maintains a Guest House for visiting scientists and experts; during the year under review practically all guests from abroad preferred to stay in the Institute, as this not only eliminated long journeys from the city but gave them the opportunity of more intimate social contacts within the Institute. As it became necessary to receive a very large number of foreign guests at the time of the 25th Anniversary in 1956, some hutments were erected to provide accommodation. During the year under review accommodation was provided for (*a*) 140 foreign and (*b*) 36 Indian guests with a total of (*c*) 549 and (*d*) 2636 guest-days respectively, or for (*e*) 176 guests for 3185 guest-days altogether. The Guest House made arrangements for showing the visitors round Calcutta and its environments and visits to cultural shows (dance, drama, music, cinema, etc.).

9.6. Canteen

The Institute Canteen continued to serve subsidized tiffin and provide meals and refreshments broadly on a cost basis. It also took charge of catering arrangements for the numerous Institute receptions and functions. About 1400 workers availed themselves of the subsidized tiffin daily and about 100 workers had other daily meals. The canteen had a staff of 35 and a total turnover of Rs. 1.9 lakhs during the year under review.

9.7. Cultural and Other Activities of Workers

A considerable expansion of the cultural and other activities of workers was noticed during the year. The ISI Workers' Club at Baranagar, the ISI Field Workers' Club with headquarters in Calcutta and the Salboni Club at Giridih organized usual sport-meetings and recreations and cultural activities. A journal under the name of *Lekhan* was brought out by the ISI Workers' Club at Baranagar, which also runs a fairly popular Cooperative Credit Society. The representative workers also maintained contact with the administration to draw attention to the grievances of workers and to offer suggestions for the improvement of conditions of work.

9.8. Samvadadhvam

The second issue of the Institute Workers' House Magazine, "Samvadadhvam" came out in October 1956 with Bengali and Hindi supplements. A special anniversary number of the journal was published in March 1957.

10. EXTERNAL ACTIVITIES

1. Professor P. C. Mahalanobis left India on 8 April 1956 accompanied by Mrs. Mahalanobis to attend the Ninth Session of the United Nations Statistical Commission held in New York (16 April to 3 May 1956) as Chairman of the Commission. At the invitation of H. E. the Governor of Puerto Rico, Professor and Mrs. Mahalanobis went to San Juan for a visit of one week beginning on 6 May 1956 to see what was called "Operation Bootstrap" and which was intended to celebrate the turning point in the progress of the national economy, when the contribution of the industrial sector overtook that of the agricultural sector to the national income. Professor Mahalanobis visited the Planning Board and the

Department of Economic Development, and had discussions with the members of the staff of these agencies and also with the Governor and some of the members of the Government. He also spoke on Indian planning at an economic seminar. On his return to USA, he went to Washington, where he met a number of leading economists and businessmen and also visited and had discussions at the Bureau of Standards and the Bureau of Census. He gave a lecture on planning in India and had some discussions at the Yale University and the Cowles Commission for Research in Economics.

2. Professor Mahalanobis gave two public lectures on planning in India, a subject in which considerable interest is taken in the USA— one at the New India House in New York, arranged by the Indian Embassy and another at the Economists' Club at the United Nations Headquarters. He also participated in a discussion meeting on the subject at the Columbia University held under the auspices of the Faculty of Economics. He received invitations from a number of other places in the USA, but the whole programme had to be cancelled, as he was taken ill. His illness made it necessary also to cancel his scheduled trips to Japan (where a delegation from the Indian Statistical Institute was waiting for him) and from there to China. Professor Mahalanobis left USA on 4 July, and after a short halt in London, returned to India on 12 July 1956.

3. A delegation from the ISI consisting of Sri Pitambar Pant, Sri M. Mukherjee and Sri R. Natarajan visited China and Japan during summer 1956. The delegation studied various aspects of planning in these countries and returned with a great deal of valuable materials. Full details are given in Appendix 15.

4. Dr. C. R. Rao attended the Third All Union Mathematical Conference in Moscow from 24 June to 5 July 1956. He presented two papers at the Conference in Moscow— one on Estimation and another on the Analysis of Growth Curves; and also visited several institutes of statistical education in the USSR.

5. Sri D. B. Lahiri attended, as a member of the Indian Delegation, the Fourth Regional Conference of Statisticians organized by the Economic Commission for Asia and the Far East at Bangkok from 26 March to 6 April 1956. He made many contributions to discussions on census problems and methods.

6. The following members of the staff of the Institute attended the Science Congress Session held in January 1957 at Calcutta and presented papers and/or participated in the discussions: Debabrata Basu, Indra Mohan Chakravarty, Sibabrata Chatterjee, Bhupendra Chandra Das, Bhagabat Das Gupta, A. E. Harper, Radha Govinda Laha, Sivkumar Mitra, Samarendra Kumar Mitra, S. Janardan Poti, C. S. Ramakrishnan, C. R. Rao, Jogabrata Roy, Subodh Kumar Roy Choudhury, Satya Prakash Sangal, Tapas Kumar Sen and V. S. Varadarajan.

7. S. Naqvi, N. K. Chatterjee, Sitangshu Bhattacharyya and Chitta Mitra attended the Indian Economic Conference and Indian Agricultural Conference in Cuttack held in December 1956.

8. S. Sen Gupta, A. Chakravarti, A. Biswas, Ajit Das Gupta and N. C. Ghosh attended the Preliminary Conference on Research on National Income held in New Delhi in February 1957.

9. Dr. Rhea S. Das served as a personnel selection consultant to Hindustan Aircraft (Private) Ltd., Bangalore.

10. Dr. C. R. Rao served as a member of Board of Studies in Statistics in a number of universities. Many other members of the Institute staff served on various Committees and actively participated in scientific and technical activities organized by other institutions and agencies.

11. VISITING SCIENTISTS

1. As in previous years, the Institute was fortunate in having a number of distinguished scientists from various countries as visitors. Some of them came as visiting professors to work in the Institute for several months, while others came for shorter periods of work participating sometimes in lecture or seminar activities. In addition, some 25 scientists came from abroad to participate in the lectures and seminars arranged in connexion with the 25th Anniversary Celebrations of the Institute held in the winter of 1956–57.

2. An account of the work of the visiting scientists who stayed for a relatively long period is given below.

1 & 2. **Professor J. B. S. Haldane** and **Mrs. Haldane (Dr. Helen Spurway)**, *University College, London* (July to September 1956). Professor Haldane delivered eight lectures on biometry, three lectures on 'genetical effects to be expected from artificial radio-activity,' and one on 'occupational mortality.' These lectures were attended, besides workers and students of the Institute, by representatives from other scientific institutions as well as by members of the public interested in the subjects. He also conducted a series of experiments on the behaviour of air-breathing fish in collaboration with Mrs. Haldane as noted earlier in the report. Arrangements were also made for Professor and Mrs. Haldane to visit a number of scientific institutions and universities, where they delivered lectures on subjects of general scientific and cultural interest as also on topics connected with statistical methods in Biology. The institutions visited by them were: Indian Council of Agricultural Research, New Delhi; Indian Central Jute Committee, Calcutta; Central Leather Research Institute, Madras; Delhi School of Economics, Delhi; Indian Institute of Technology, Kharagpur; Central Building Research Institute, Roorkee; Indian Institute of Science, Bangalore and the Universities of Allahabad, Andhra, Annamalai, Mysore, Nagpur, Poona and Bombay.

3. **Professor G. A. Barnard**, *Imperial College of Science and Technology, London*, (September to October 1956). He delivered a course of lectures on statistical inference with special reference to current controversies and industrial applications of statistics. He worked with the SQC Unit in Calcutta and took an active part in the SQC Conference held on 1 October 1956 and drew up a report on the basis of discussions. He went to Delhi and Bombay on short visits and gave lectures there on SQC methods and their applications in India. At Bombay he also met the representatives of industry for discussion.

4. **Dr. R. L. Ackoff**, *Group Director, Operations Research, Case Institute of Technology, Ohio, USA* (January to March 1957). As already mentioned he gave a course of lectures on operations research.

5. Dr. Amar G. Bose, *Massachusetts Institute of Technology, USA*, came to the Institute with a Fullbright Scholarship (September and October 1956). He delivered a course of lectures on non-linear system characterization, filtering and prediction.

6. Professor Simon Kuznets, *Johns Hopkins University, USA* (December 1956 to February 1957) came to India accompanied by his wife, as guests of the Institute. Apart from participating in various discussions, he made a critical evaluation of the present position of national income estimation and prepared a report on the subject, a copy of which has been submitted to the Government. He took a leading part in organizing the Preliminary Conference on Research on National Income held in New Delhi in February 1957. He also gave some lectures and had discussions in Delhi. He attended the Indian Economic Conference held in Cuttack on 26-28 December 1956.

7. Mr. Alvin Mayne, *Planning Board, Puerto Rico* (February to April 1957). He worked mostly on the organization of planning activities of the ISI, visiting in this connexion the offices of the Institute in Delhi. Besides, he gave a number of lectures and participated in discussions and meetings there on Planning.

8 & 9. Mr. A. F. Zelinovsky, *Deputy Chairman, Planning Commission* and Mr. A. I. Polosukhine, *Deputy Chief of Machine Building Department and Member of the State Economic Commission, USSR* (November 1956 to January 1957). They helped in the work of planning, particularly in the field of heavy machine building industry and long-range planning. They stayed mostly in Calcutta and Delhi and also paid a short visit to Bombay in connexion with their work.

10. Dr. J. P. Miller, *Yale University, USA* (December 1956 and January 1957). Dr. Miller's itinerary in India included a visit to Delhi, where he had talks with ministers, members of the Planning Commission and Government officials. He also paid visits to Madras, Bangalore, Mysore and Bombay, where he had discussions with prominent people in the field of industry, administration and finance. Both at Calcutta and Bangalore Dr. Miller delivered some lectures on 'industrial administration and management.'

11 & 12. Sir Robert Hall, *Economic Adviser to the Treasury of the Government of UK* and Lady Hall, *Staff member of the University of Oxford* (October to November 1956) participated in a number of discussions on planning in Calcutta and New Delhi. Sir Robert gave a few lectures on planning; he also prepared a memorandum on the possible role of ISI in the field of applied economic research necessary for governmental purposes.

13. Professor J. S. Neyman, *Statistical Laboratory of California University*, stayed in the Institute from December 1956 to February 1957.

14 & 15. Mr. Leo Huberman, *Editor, Monthly Review, New York, USA* and Mrs. Huberman also visited the Institute (January to March 1957).

Foreign Lecturers

The following scientists, who came to the Institute on brief visits, gave lectures and took part in discussions in the Institute.

1. Mr. Kailas Doctor, *International Labour Organization, Geneva* (February 1957). Lecture: Statistical work of the ILO.

2. PROFESSOR G. L. STEBBINS, *Chairman of the Department of Genetics in the University of California* (January 1957). Lecture: Non-biometric analysis of natural variation in plants.

3. DR. A. T. BHARUCHA REID, *University of Oregon, USA* (August 1956). Lectures : (*i*) Estimation of exponential parameter associated with Murkoff's process, (*ii*) Comparison of populations whose growth can be described by a branching stochastic process and random elements in Orlion spaces.

4. DR. NATHAN KEYFITZ, *Dominion Bureau of Canada* (July 1956). Lecture: Population sampling in Canada.

5. MR. L. FISCHMAN, *Puerto Rico* (12 December 1956 to 12 February 1957). He participated in the work on planning and attended the preliminary conference on Research in National Income in New Delhi, apart from delivering a few lectures.

INDIAN LECTURERS

The following lectures were delivered by Indian statisticians and economists.

1. PROFESSOR K. V. KRISHNA SASTRY, *University of Madras* (May 1956). Lecture: Consumer's demand for products of cottage and small scale industry.

2. PROFESSOR S. K. EKAMBARAM, *University of Mysore* (August 1956). Lecture: Control charts in production.

3. PROFESSOR J. K. PARANJAPE, *Sydenham College of Commerce and Economics, Bombay* (November 1956). Lecture: State enterprises in India and Planning in Poland.

4. SRI JAGJIT SINGH, *Railway Board* (February 1957). Lecture: Foundations of Probability.

5. DR. K. S. BANERJEE, *State Statistical Bureau, West Bengal* (August 1956), delivered a course of six lectures on Index number.

The above list does not include the large number of lectures and seminars arranged in connexion with the 25th Anniversary Celebrations of the Institute. A list of these lectures is given in Appendix 2G.

12. GENERAL ADMINISTRATION

1. *Membership :* The position of membership of the Institute during the year under review was as follows :

	new members enrolled during the year	*total number at the end of the year*
Honorary members	nil	8
Life members	nil	47
Ordinary members	33	179*
Sessional members	nil	9
Student members	nil	21

2. *Council :* The names of the members of the Council are given in Appendix 16. The Council held 16 meetings during the year. Among the important matters considered (6 April 1956): bringing into force amended rules 11 and 14 (5 May 1957); appointment of

---

* The actual number of ordinary members on the roll was 531 on 31 March 1957 but of these only 179 were effective members.

a panel of auditors (13 June 1956); arrangements of 25th anniversary celebrations, giving effect to amended rule 11A (25 August 1956); amendment of provident fund rules by making provision for sanction of loans for housing purposes, increase of dearness allowance (28 August 1956); consideration of the Indian Statistical Institute Bill, 1956 (14 and 22 December 1956); introduction of new pay-scale (18 December 1956); changes in some pay-scales, revision of leave rules of workers on a daily wage basis and field workers, revision of rules guiding employees' and employers' contributions to the provident fund (4 February 1957); renewal of agreement with Statistical Publishing Society (29 March 1957). The meeting of 22 December was held jointly with a meeting of the Governing Body. This meeting discussed in detail the proposed legislation for declaring the Institute as an institution of national importance and set up a Committee consisting of Sri C. D. Deshmukh, Sir Shri Ram, Sir D. N. Mitra, Professor D. R. Gadgil, Professor K. B. Madhava and Professor P. C. Mahalanobis to examine the proposed legislation and submit their report to the Council.

3. *Governing Body :* The Governing Body of the Research and Training School met at New Delhi on 9 October 1956 and at Calcutta on 22 December 1956. Sri C. D. Deshmukh, President, presided over both the meetings. The Finance Committee of the Governing Body met on 9 October 1956 at New Delhi. (The names of the members of the Governing Body, the Finance Committee and other Committees set up by the Council are given in Appendices 17, 18 and 19).

4. *Finance Committee :* The Finance Committee of the Council met on 4 February 1957 and took important decisions in respect of changes in certain pay-scales, changes in rules guiding leave and leave salary on medical certificates and changes in employees' and employers' contribution to Provident Fund.

5. *General Meeting :* At a General Meeting of the Indian Statistical Institute, held on 20 April 1957, certain amendments to the rules of the Institute were approved. The Annual General Meeting of the Institute was held on 29 October 1956.

6. *Board of Management :* A Board of Management was appointed on 28 August 1956 with Professor Mahalanobis as Chairman, Sri Sidhartha Banerjee as member-secretary and Sri D. B. Lahiri, Sri M. L. Ganguli, Sri Samar K. Mitra and Dr. C. R. Rao as members. The Board was entrusted with the general management of most of the operative departments and some common services sections. During the period under review the Board held 17 meetings and took decisions on 116 items.

7. *Cost Accounting Section :* This section was concerned with evaluation of output records of operational items, supplied by primary workers of project divisions, in terms of equivalent standard hours. Besides the usual primary workers, technicians, supervisors and others in senior cadres in the NSS regularly supplied their weekly diary records to this section for analysis and for the working out of standards for various technical jobs. As the incentive bonus scheme was practically withdrawn during the year under review, the Board of Standard also ceased to function.

8. *Calcutta City Office :* The Calcutta SQC Unit had its office during the year at 9B Esplanade in the city. This office was also used for general administration work and for meetings and lectures of the Institute. The Institute has two other offices in the city, one at 294/2/1 Upper Circular Road, Calcutta-9 and the other at 210 Cornwallis Street, Calcutta-6.

9. *Headquarters :* During the year under review hutments (with asbestos roofing) containing 20,000 sq. ft. of internal floor area were constructed at 202 and 203 B. T. Road. In addition, two cottages (also with asbestos roofing) with a floor space of about 8,800 sq. ft. were constructed in the northern and eastern parts of 202 B. T. Road to provide accommodation for the guests who attended the 25th Anniversary Celebrations in December 1956 and later on for visiting scientists, trainees of the International Statistical Education Centre and officers on deputation for training or work. Certain other temporary structures were put up and extensions made to existing structures to meet a variety of purposes including the storing and maintenance of Soviet equipment received through the UNTAA.

10. *Giridih Office :* A fair amount of construction and repair work was done under the Giridih office during the year under review. As already mentioned, a Health Home was opened at Giridih in March 1956. Various experiments and surveys were carried out under the direction of the Giridih office, further details of which are given in Section 4.

11. *Delhi Office :* The office at Delhi continued to function at 8 King George Avenue as a link between the Institute and the Cabinet Secretariat, the Central Statistical Organization, the Planning Commission and various other Government departments and other agencies.

12. *Bombay Office :* The Bombay Office helped the Bombay branch of the Institute in carrying out the work of the 11th round of the NSS in the Bombay city. The office also helped in conducting professional examinations of the Institute at the Bombay centre. A series of lectures were also organized as mentioned subsequently in the report on the Bombay branch of the Institute. The activities of the SQC, Bombay Unit have already been mentioned.

13. *Field Branch :* The field branch of the Institute, the office of which is located at 294/2/1 Upper Circular Road, Calcutta-9, took up a series of experimental surveys over and above the field work in connexion with the National Sample Survey in West Bengal. Some of these special surveys have already been covered in the sections on project work.

14. *Distribution of staff at different centres :* The following table shows the distribution of the staff at Baranagar, Calcutta, Giridih, Delhi, Bombay and Bangalore as on 28 February 1957 as compared with the figures at the end of the previous (financial) year. Figures are also shown for the field branch maintained under the direct control of the Institute for sample surveys and special enquiries.

centre	general workers		subordinate staff		total	
	1956	1957	1956	1957	1956	1957
Calcutta Headquarters (Baranagar)	756	907	290	390	1046	1297
Calcutta City	9	22	9	7	18	29
Giridih	100	100	29	33	129	133
Delhi	28	38	5	8	33	46
Bombay (SQC)	10	14	1	4	11	18
Bangalore (SQC)	8	17	1	3	9	20
IMRUP (Bangalore)	1	24	–	4	1	28
	912	1122	335	449	1247	1571
Statistical Field Branch including Bombay NSS	132	145	41	41	173	186
total	1044	1267	376	490	1420	1757

15. *Changes in senior staff :* The staff of the Institute was strengthened during the year by the addition of the following persons: Sri D. P. Basu joined IMRUP on 2 April 1956; R. B. Banerjee joined the Electronics Section on 1 August 1956; Dr. Raghuraj Bahadur joined the RTS on 1 December 1956; Sri Teru Sasaki joined the Planning Division on 12 December 1956; Sri K. V. Ramachandra joined the IMRUP on 2 January 1957. Dr. Des Raj went on leave in September 1956 to take up an assignment under the American University of Beirut. Sri Ranjan Som was given leave for one year to accept a post of Statistician under WHO at Colombo. Sri K. Naqvi of Aligarh University worked in the Institute for about 9 months. Dr. Asoke Rudra was given leave for two years for higher studies in France.

The following persons left the Institute during the year: Sri Mohi Mohan Mukherjee left in August 1956 to join Philips Electrical Co. (India) Ltd.; Dr. P. K. Gopalkrishnan left in September 1956 to join Delhi University; Sri I. R. K. Sarma left in December 1956 to join the National Council of Applied Economic Research.

16. *Obituary :* (1) The sudden and unexpected death of Professor J. M. Sen during this year was a great blow to the Institute. The late Professor Sen had a heart attack while coming to the Institute on 29 August 1956 and died in hospital the next day. He had been a member of the Institute for six years and a member of the Council for five years. At the time of his death, he was Secretary of the Examinations Committee of the Institute, an office which he had held for a year and a half. He had served the Institute with great devotion and initiative and earned the respect of everybody who came in contact with him. The Council of the Institute recorded its deep sense of sorrow at his sudden passing away in an appropriately worded resolution at its meeting held on 5 October 1956.

(2) The Institute also suffered another grievous loss in the sudden and untimely death on 23 February 1957 of Professor J. B. Saunders, who had joined the Institute as, a Professor of Economics on 23 January 1957. The Council mourned his death at its meeting held on 2 March 1957.

17. *Sankhyā : The Indian Journal of Statistics :* Parts 3 and 4 of Volume 16 and all the 4 parts of Volume 17 of Sankhyā containing in all 40 papers and 2 reports were published during the year. There was a separate issue containing papers on design of experiments. In addition, there were papers on sample surveys, time series analysis, stochastic processes, etc., as well as some important contributions on the Second Five Year Plan of India. Part 1 of Volume 18 of the journal was in press at the time of writing this report. There has been a steady increase in the demand for the journal and a number of new subscribers were enrolled during the year.

13. BRANCHES

13.1. Bombay Branch

1. *Membership :* At the close of the year under review, there were 44 associated members, 3 ordinary members and 1 life member of the branch. The names of office-bearers of the branch are given in Appendix 18.

2. *Sample Surveys :* The branch carried out the field work in connexion with the eleventh round of the NSS in Bombay City from August 1956 to February 1957. The

field work of the Pilot Enquiry into the household purchase of products of factories and cottage and small-scale industries in Bombay was completed during the year. The branch also published a report on the "Survey into the economic conditions of middle class families in Bombay City."

3. *Institute Examinations :* The Statistician's Diploma Examination was conducted at the Bombay Centre in September 1956.

4. The Rt. Hon. John Strachey, M.P., Professor Oscar Lange, Mr. A. Zelinovsky, Mr. A. Polosukhine, Professor J. Neyman, Mr. J. P. Miller, Sir Robert and Lady Hall, Professor and Mrs. Simon Kuznets and Mr. and Mrs. Leo Huberman were among the distinguished visitors to the branch during the year.

5. The activities of the SQC Unit in Bombay have already been described in Section 6.

13.2. Mysore State Branch

1. *Membership :* At the close of the year, there were 29 ordinary members, 2 sessional members and 2 life members bringing the total membership of the branch to 33. The names of the office-bearers of the branch are given in Appendix 18.

2. *Statistical Surveys :* A pilot sample survey was undertaken on "Job satisfaction among workers in industrial concerns in Bangalore;" the scheme for another pilot sample survey on "Variation of output efficiency during a shift and in different shifts" was drawn up during the period under review.

3. *Institute Examinations :* Statistical Field Survey Certificate Examination was conducted at the Centre in September 1956.

4. *Training classes and lectures :* In response to requests from various industrial establishments, the branch conducted a seven-week course on "Methods of statistics for quality control worker." Lectures were organized on different subjects. Professor J. Neyman gave a talk on "Probability, Statistics and Inductive Behaviour."

13.3. Aligarh Branch

1. The activity of the branch suffered a set-back as a result of the continued ill health of its President, Professor D. P. Mukherjee. Dr. Zakir Hussain, who as first President of the branch had taken active interest in its work, relinquished the office of Vice-Chancellor of the Aligarh University during the year, a fact which affected the activities of the branch to a certain extent. The names of office-bearers are given in Appendix 18.

2. The branch undertook an intensive survey on the process and magnitude of income generation in two villages. A study of under-employment in two villages was also undertaken and a report submitted to the parent body. A survey of local industries in Aligarh was also carried out and a brief report prepared.

Indian Statistical Institute : Receipts and Payments Account

To Receipts	Rs.	As.	P.	Rs.	As.	P.
1. Opening Balance :						
i) Cash in hand and at Banks with Central Office	36,856	15	8			
ii) With branches and sub-offices	26,774	5	0			
iii) Unadjusted suspense of 1955-56	75,010	1	2			
				1,38,641	5	10
2. Advances recoverable :						
i) From Capital Expenditure account as per last account	1,49,682	4	0			
ii) Educational and house building loan etc. due from staff as per last account	14,984	1	0			
				1,64,666	5	0
3. Membership subscription				4,084	9	0
4. Fees for tuition and training				12,702	8	0
5. Examination fees and other receipts				27,380	13	6
6. Receipt by recovery of expenses incurred for training				4,962	15	6
7. Seminar fees received by IMRUP				600	0	0
8. S.Q.C. membership and training fees including service charges				97,644	6	0
9. Block and Project grants from the Government of India for—						
i) Research, Training and General purposes ..	8,07,200	0	0			
ii) International Statistical Education Centre ..	92,000	0	0			
iii) Statistical Quality Control Sector	1,24,000	0	0			
iv) Electronic Laboratory and Computing machine Sector	2,42,000	0	0			
v) Economic Wing	2,65,500	0	0			
vi) Operational Research Sector (for Planning) ..	3,49,000	0	0			
vii) Industrial and Management Research Unit for Planning	1,74,000	0	0			
viii) Utilisation of the UNTAA (Soviet) equipment ..	2,53,000	0	0			
ix) Multipurpose National Sample Surveys ..	42,54,500	0	0			
				65,61,200	0	0
	Carried over ..			70,11,882	14	10

(current expenditure) for the year ending 31st March 1957

By Payments	Rs.	As.	P.	Rs.	As.	P.
1. (i) Salary, dearness allowance, honorarium etc. ..	34,45,569	10	9			
(ii) Employer's contribution to workers' Provident Fund	1,44,060	7	6			
(iii) Overtime allowances	1,31,057	13	6			
				37,20,687	15	9
2. Contribution to Leave Salary Fund				2,00,000	0	0
3. Contribution to Gratuity Fund				1,50,000	0	0
4. Travelling allowances				1,70,147	10	3
5. Visiting Professors, Fellows, Foreign experts and scientists etc.:						
i) S.Q.C. Sector	23,559	7	0			
ii) O.R.U. Sector	94,939	13	6			
iii) UNTAA Sector	34,462	0	6			
iv) N.S.S. Sector	81,451	11	0			
	2,34,413	0	0			
v) Research and Training Sector (transfer to fund account)	60,000	0	0			
				2,94,413	0	0
6. Scholarships, stipends and assistance to trainees of the R. & T. School (transfer to fund account) ..				1,50,000	0	0
7. Disbursement of fellowship allowance to ISEC trainees under Colombo Plan				63,307	8	6
8. Prizes to workers for initiative etc.				24,000	0	0
9. Machine Tabulation expenses:						
a) Hire and maintenance of tabulating equipment, key punches and verifiers including freight, transport, electricity charges etc.	7,73,904	6	3			
b) Cost of cards, cabinets etc.	2,57,693	6	3			
c) Payments to IBM, BTM, Powers-Samas and Gokhale Institute of Politics and Economics, Poona for tabulation of N.S.S. data	73,660	15	0			
				11,05,258	11	6
10. Printing and publication (including paper for printing)				1,01,479	15	0
11. Society type activities				30,749	2	5
12. Examination expenses				24,331	1	6
13. Books and Journals (including cost of binding etc.) ..				75,117	2	1
14. Translation Unit set up in Japan for planning purposes				24,468	11	0
15. Workshop, photo and microfilm				18,101	12	3
16. Laboratory and workshop stores, tools and minor accessories etc.				63,744	14	3
17. Repairs and replacement of machineries, equipment, accessories, furniture and fittings etc.				88,742	10	3
18. Stationeries and consumable stores				1,26,336	3	6
19. Auditor's fees				3,600	0	0
20. Bank charges and interest				12,848	15	3
21. Telephone, postage, telegram, advertisement, freight charges etc.				75,709	4	9
Carried over ..				65,23,044	10	3

Indian Statistical Institute : Receipts and Payments Account

To Receipts (Contd.)

	Rs.	As.	P.	Rs.	As.	P.
				B.F. 70,11,882	14	10
10. Amount received from the United Nations for study on existing information on population in relation to development planning in India				2,381	0	0
11. Grant received from the Government of India to meet the expenditure for 25th Anniversary celebrations of the Institute				2,00,000	0	0
12. Funds received from the Government of India, Ministry of Finance for disbursement to the trainees at ISEC selected as fellows under the Technical Co-operation Scheme, Colombo Plan				67,180	0	0
13. Miscellaneous receipts :						
a) Donations and contributions	3,396	10	3			
b) Sale proceeds of waste paper, cards etc. ..	3,561	9	3			
				6,958	3	6
14. Receipts from produce of Giridih experimental farm .				3,690	0	9
15. Deposit accounts				6,988	6	0
16. Outstanding liabilities for goods and services obtained				1,10,640	13	3
17. Temporary loan taken from funded accounts ..				3,40,000	0	0
18. Temporary loan taken from Prof. P. C. Mahalanobis .				50,000	0	0
				Rs. 77,99,721	6	4

6, HASTINGS STREET, CALCUTTA,
The 25th September, 1957.

(current expenditure) for the year ending 31st March 1957

By Payments (Contd.)

	Rs.	As.	P.	Rs.	As.	P.
B.F.				65,23,044	10	3
22. Other miscellaneous expenses including hot weather contigencies etc.				45,236	3	9
23. Electricity charges				35,621	0	6
24. Rent, rates and taxes (including those of field, camp and outstation offices)				1,71,291	3	3
25. Repairs and maintenance of land and buildings including petty constructions				1,43,117	7	9
26. Transport				54,307	8	6
27. Workers' welfare and amenities				1,05,014	14	0
28. Development at Director's discretion				3,672	8	0
29. Transport, freight and installation charges of Soviet (UNTAA) equipment				20,183	1	9
30. Expenditure for 25th Anniversary celebrations of the Institute				,137,267	9	9
31. Expenditure on account of experimental farming at Giridih				3,688	10	9
32. Repayment of outstanding liabilities (as per debits shown in last account)				1,48,884	6	6
33. Repayment of old deposit (as per last account) ..				5,514	5	3
34. Loans to staff for educational and house building purposes etc.				22,241	4	6
35. Amount under suspense pending final adjustment :						
a) part payment to contractors and suppliers against work done or goods supplied or ordered, t.a. advances and other advances awaiting adjustment ..	67,667	12	9			
b) for passage, custom deposits etc. of Visiting Professors and Foreign Experts	48,249	0	3			
c) Advances to small industries experimental unit for operational expenses	25,252	9	3			
i) run directly by I.S.I.						
ii) run through Prof. N. K. Bose	4,900	0	0			
				1,46,069	6	3
36. Temporary advances to Capital Expenditure :						
a) as per last account	1,49,682	4	0			
less amount adjusted during the year ..	27,463	12	4			
	1,22,218	7	8			
b) fresh advance during the year	50,000	0	0			
				1,72,218	7	8
37. Closing Balance :						
i) Cash in hand and at Bank with Central Office ..	26,616	14	11			
ii) With branches and sub-offices	35,731	11	0			
				62,348	9	11
				Rs. 77,99,721	6	4

Examined and found correct
Sd/- P. C. Nandi & Co.
Chartered Accountants & Auditors.

PART 3: APPENDICES

Appendix 1: A brief history of the Institute

1. The beginnings of the Indian Statistical Institute were modest enough: a solitary computer working part-time and a total current expenditure of rather less than Rs. 250—that was all, when it started in April 1932. In September 1956, the Institute had on its roll 1674 regular workers, who were backed by an annual budget allocation of Rs. 52,00,000. Impressive as these figures are, they convey little idea of the nature of the work that is being carried on by the Institute, its wide range, its complex character and the specialization at high scientific levels, that is necessary for carrying out investigations which have an intimate bearing on the life of the nation.

2. A long period of preparatory activity, going back to the early years of the first world war, preceded the formal inauguration of the Institute. Prasanta Chandra Mahalanobis was back in India in 1915 on a short holiday after he had been awarded a senior scholarship of King's College, Cambridge, for research in Physics. Just before his departure from Cambridge, his tutor, W. H. Macaulay, had drawn his attention to the *Biometrika* and Karl Pearson's *Biometric Tables*, copies of which he brought with him to India and which he studied with eager interest while waiting to return to Cambridge for his Physical researches. But he did not return to Cambridge. Accepting the post of a professor of Physics in the Presidency College, Calcutta, he decided to stay back in India and carry on his statistical studies alongside of his work as a teacher of Physics. In Calcutta, he came into intimate contact with the eminent scholar and philosopher, Brajendra Nath Seal, who had a full appreciation of the importance of modern statistical methods and encouraged Mahalanobis to take up this subject as his life's work.

3. Urged by the great savant, young Mahalanobis entered upon his statistical career with gusto. Not content to confine his interest in statistics to merely academic studies, he was soon applying statistical methods for the solution of problems in anthropometry, which led to the formulation of the Generalized Distance (D^2-statistics). In meteorology, he found ample scope for trying out his new-found tool, enabling him to locate, at a height of about four kilometers above sea level, the region of highest control for changes in meteorological conditions on the surface of the earth; later, this result was confirmed by Franz Baur from physical considerations.

4. More far-reaching from the point of view of national welfare were the results of extensive investigations into rainfall and floods in North Bengal, Orissa and West Bengal. These led not only to suggestions for really effective flood control but also supplied in later years some of the basic calculations for two great river valley schemes, namely, the Damodar Valley and the Hirakud Projects.

5. Meanwhile, Mahalanobis had gathered around him several part-time assistants and a group of young research workers to form the nucleus of the Statistical Laboratory, which was located in the Presidency College, Calcutta. Amongst this pioneer band were Sudhir Kumar Banerjee, Subhendu Sekhar Bose, Nistaran Chakraburty and Jaladhar Sarma, the only one amongst the four who is still serving the Institute. Sudhir Kumar who became the Chief Computer and Subhendu Sekhar, who was the leading spirit amongst the workers of the Institute, have both been cut off by untimely death in the prime of their careers.

6. The Laboratory also had been gradually developing. By 1930 it owned about Rs. 25,000 worth of books, calculating machines and other equipment: and had undertaken

some enquiries on behalf of the Government departments and private concerns. The first official recognition came in July 1931 with an annual grant of Rs. 2,500 from the Imperial (now Indian) Council of Agricultural Research for Statistical investigations relating to agriculture.

7. The year 1931 marks another milestone in the progress of statistical organization in India. At a meeting convened by P. N. Banerjee (the then Minto Professor of Economics), N. R. Sen (Ghosh Professor of Applied Mathematics) and P. C. Mahalanobis (Professor of Physics, Presidency College) and held on 17 December 1931, with Sir Rajendra Nath Mookerjee in the Chair, a resolution was unanimously adopted to establish the Indian Statistical Institute. The constitution was approved at the end of February and the Institute was formally registered as a non-profit society in April 1932. Branches were established quite early in Mysore, Poona, Bombay, Madras, Lahore, Banaras, Lucknow and Delhi: and the Statistical Laboratory in Calcutta continued to function as the active nucleus of the Institute.

8. This was how the Indian Statistical Institute was formally organized. The first general grant for research and training came in 1935 in the shape of Rs. 5,000 a year from the Government of India. A big development occurred when the Institute took up in 1937 a five-year project, on an expanding scale, to develop a sampling method for estimating the acreage and out-turn of the jute crop in Bengal with the required accuracy at a reasonable cost. Since then, the Institute has been an important centre of sample surveys in the whole world. From this time, the larger part of the income of the Institute began to be derived as contract grants for applied projects and enquiries.

9. The training of officers deputed by Central and State Governments had started from 1932. In the earlier years such training was more or less on an individual basis: and over 150 individuals received such training at the Institute between 1932 and 1939. From 1939, the Institute started organized courses of instructions and the training courses gradually developed into the present Research and Training School at a post-graduate level.

10. Examinations for the award of certificates for computers and diploma for statisticians were started in April 1938, and served a real need, and have become quite popular and are held all over India every year.

11. To offer facilities for the publication of research papers in India, it was decided in 1933 to start *Sankhyā*: The Indian Journal of Statistics, as the official organ. On the initiative of the Council of the Institute, arrangements were also made to establish an associated non-profit institution, the Statistical Publishing Society, to maintain the Eka Press and publish *Sankhyā*, which soon became a journal of international recognition.

12. The Indian Statistical Institute had for some time been pleading for a separate statistical section to the Indian Science Congress. These efforts, however, did not bear any fruit and so in 1938 the Institute organized on its own the First Session of The Indian Statistical Conference which continued for several years. The Indian Science Congress agreed to start an independent section for Statistics from 1946.

13. The Institute's research and project work as well as its promotional activities led to world-wide contacts, earned for it international recognition and helped in enlisting active association with it of distinguished foreign statisticians who came out to India to work and lecture in the Institute as visiting professors. The first to come was Professor

(now Sir) Ronald A. Fisher, who paid his maiden visit in 1938, repeating it in 1945, 1951, 1954 and 1956. Other visits have followed and, except for an unavoidable break during the Second World War, a stream of distinguished foreign scholars, statisticians as well as specialists in other fields of knowledge, have responded to the invitation of the Institute to come here and work as research workers and guest lecturers.

14. The Institute has thus become a real international centre of research as well as a forum for scholars and scientists from all over the world to work and hold discussions. Mention may be made of a few amongst these distinguished visitors, such as, G. A. Barnard, R. A. Fisher, Richard Goodwin, J. B. S. Haldane, Sir Robert Hall, Nicholas Kaldor, J. R. N. Stone, John Strachey, F. Yates (UK), Russel L. Ackoff, Samuel Alexander, Paul A. Baran, J. K. Galbraith, Morris H. Hansen, Harold Hotteling, W. Hurwitz, Leon H. Keyserling, Simon Kuznets, Alvin Mayne, John Perry Miller, J. S. Neyman, E. R. Ott, Walter A. Shewhart, Abraham Wald, S. S. Wilks, Norbert Weiner (USA); N. N. Bogolyubov, D. D. Degtyar, V. A. Ditkin, A. I. Ezhov, Y. V. Limmik, I. Y. Pisarev, A. I. Polosukhine, M. I. Rubinstein, A. F. Zelinovsky (USSR); Faiz El Khuri, (Beirut); Nathan Keyfitz (Canada); Wang Szu Hua, Hwang Chien To, Tai Shin Kwang (China); C. Bettelheim (France); Tosio Kitagawa, Motosaburo Masuyama, R. Minobe (Japan); J. A. Links, E. Lunenberg and J. Timbergen (Netherlands); Ragnar Frisch (Norway); Q. M. Hussain, M. Ziauddin (Pakistan); Oskar Lange (Poland); H. Wold (Sweden); Arthur Linder (Switzerland); M. C. Athiporn P. Ksemasri (Thailand).

15. After the Second World War, the project side of the Institute developed rapidly to meet the increasing need for statistics of the Government. The Sample Survey in Bengal was extended in 1943 to cover both the area sown and the total yield of jute, rice and other important agricultural crops throughout the year. The tabulation on the basis of the two percent *Y*-sample of the 1941 Population Census started in 1944-45, and various other socio-economic surveys were undertaken.

16. The post-war years have witnessed a rapid expansion and many new developments. The question of stabilization of the Institute had been under consideration of the Government of India since 1938. Sri Chintaman Deshmukh, who has been the President of the Institute since 1945, helped in solving many difficulties; and it was mainly through his efforts that the Research and Training School was established with an initial recurring grant of Rs. 4.5 lakhs from the Government of India from 1949-50. Administrative sponsorship of the Institute was transferred from the Ministry of Education to the Ministry of Finance in 1950, and to the Cabinet Secretariat of the Government of India in 1956. A Governing Body was established in 1952 to look after the affairs of the Research and Training School, subject to general coordination of the work of the Institute as a whole by the Council.

17. From about this time, the Institute began to emerge as a national institution. In 1950, the Institute started working on a vast project, namely, the design and analysis of the data of the National Sampling Survey, which is collecting comprehensive information relating to social, economic and demographic characteristics on a countrywide basis in the form of two "rounds" of survey every year covering both rural and urban areas.

18. In 1950, the Institute also helped to bring into being, in collaboration with the International Statistical Institute and under the sponsorship of the UNESCO, the International Statistical Education Centre, which is being maintained with the financial support of the Government of India. Since then, the ISEC in Calcutta has been providing statistical training to students from many Asian countries; about 202 trainees have come upto 1956.

19. Through all these years the Institute had been steadily gaining in prestige both in India and abroad. Professor Mahalanobis was elected a Fellow of the Royal Society in 1945, and acted as the Chairman of the United Nations Sub-Commission on Statistical Sampling from 1947 to 1951; he has been a member of the UN Statistical Commission since 1946 and Chiarman since 1954. Several other workers of the Institute enjoy an international reputation and some of the old members of the staff now hold important positions in other places in and outside India. The Institute acted as the host society to the International Statistical Conferences in India in 1951.

20. Since 1949 when Professor Mahalanobis became Honorary Statistical Adviser to the Central Cabinet, the Institute has been more and more closely associated with the Central Statistical Organization (CSO) and other Government agencies in New Delhi. Since 1952, the Institute is functioning as the focal centre for professional training and research and as a National Statistical and Computational Laboratory.

21. An associated non-profit institution for the development of calculating machines and scientific instruments had been established in 1943 with the approval of the Council of the Institute and a small workshop started, which, however, did not make much progress in the beginning. A small workshop was started in the Institute in 1950 for the repair and maintenance of desk calculating machines and other equipments a little later and is now in operation. An electronic laboratory was started a little later, where a small electronic analogue computer was designed and constructed in 1953. An electronic digital computer also was purchased. In 1955, arrangements were made to secure a bigger digital computer and a large number of machine tools from the USSR through the United Nations Technical Assistance Administration and further developments are in progress.

22. The Institute has been promoting the introduction of Statistical Quality Control for a long time, and it started the earliest courses in this subject in India in 1945-46. A separate SQC Section was established in 1953, which has whole-time units working at Bangalore, Bombay and Calcutta.

23. Still more significant were the developments in 1954. The Indian Statistical Institute was called upon by the Planning Commission to undertake, jointly with the CSO, work on national planning with the dual objective of solving the unemployment problem in 10 years and continuously increasing the national income as rapidly as possible. In November 1954, Prime Minister Nehru inaugurated in the Institute studies relating to planning for national development. This work is carried on with the active collaboration of the CSO, the Department of Economic Affairs (Ministry of Finance) and the Economic Division of the Planning Commission. Many of the foreign experts, who have come to the Institute, have also participated in studies and discussions on planning.

24. On the basis of the above work, a "Draft Plan-Frame" was submitted to the Prime Minister in March 1955, which was accepted in May 1955 by the National Development Council as the basis for the formulation of the Second Five Year Plan. The Institute thus became intimately associated with the work of national planning in India.

25. Since its inception, there has been a continuous expansion in diverse directions of the area of the Institute's activities. But all these developments have been promted by one common purpose, which uniforms the varied activities carried on by the different departments of the Institute: the promotion of national development through the patient collection and analysis of statistical and technical information and their utilization for policy and administrative decisions in a scientific manner.

Appendix 2 : Twenty-fifth Anniversary Celebrations

1. The Indian Statistical Institute completed 25 years of its existence in December 1956. The celebration of this event in a befitting manner was one of the most notable features of the activities of the Institute during the period covered by the present Report (April 1 1956 to March 31 1957).

2. Preparations for the celebrations were taken in hand months before the actual event, starting on 28 March 1956 with a letter addressed by the Director of the Institute to the workers outlining his ideas in regard to the celebration. A Preparatory Committee was formed soon after, and at a meeting of all the workers organized by this Committee and held on 4 April 1956 the Director referred to the 25th Anniversary as a turning point in the history of the Institute. The accent from the very beginning was on the organization of lectures, seminars, demonstrations and exhibitions in connexion with the Anniversary Celebrations designed to explain the history and practice of statistics in particular and in general, the application of science to efforts for national development in India and human welfare all over the world.

3. Such a programme depended for its success on the participation of leading statisticians as well as scholars in other fields of study from different parts of the world, and invitations were issued to many foreign scientists. All this necessitated the provision of adequate funds. One of the very first steps taken in the planning of Anniversary Celebrations was, therefore, the preparation of a budget for a special grant for the purpose and its submission to the Union Government. Sanction was obtained in due time to enable the authorities of the Institute to go ahead with the preparations.

4. The actual celebrations took place over a period of about 3 weeks starting from 17 December 1956, the two main functions being the Inaugural Meeting held on the 17th and the Anniversary Meeting held on the 22nd. A large number of guests were invited to both these meetings.

Inaugural Meeting : 17 December

5. Sri Chintaman D. Deshmukh, President of the Institute, inaugurated the celebrations at a meeting held on 17 December at 4-30 p.m. in the mango-grove. After the chanting of a vedic hymn, Sri Deshmukh read out to the large gathering, which had assembled on the occasion, the following message received from the Prime Minister, Sri Jawaharlal Nehru :

> The Indian Statistical Institute has gained a reputation for itself not only in India but in every country where statistics are studied. It has grown in recent years and has performed important functions in connexion with our planning work. Indeed, it is an essential part now of our planning organization. On the occasion of its Silver Jubilee, I send it all my good wishes.

6. Sri Deshmukh then made a short speech welcoming the guests, particularly those who had come from abroad. He described how the Institute was founded in 1931 as an outcome of the early statistical studies of Professor Mahalanobis, which were started in 1920's. He then mentioned as a curious coincidence the fact that while these early statistical studies were proceeding in Calcutta, Sir John Hubback had initiated the use of random

sampling to estimate the yield of rice per acre in Orissa. A little later, he (Deshmukh) was attracted by Hubback's method and used it to estimate agricultural production as a Settlement Officer, while Professor Mahalanobis independently used similar methods in his work on crop surveys and was much impressed by Hubback's work. It was some years later, he said, that he came to know of the existence of the Institute and when he joined the Institute's Governing Body and later became its President in 1945, he was able to render some assistance in securing financial help from the Government. Increasing contact with the Government led to the realization by the latter that the Institute should be utilized for the formulation of its policies, particularly in the field of planning.

7. Some of the old workers of the Institute then told the gathering about their experiences as workers. Professor Mahalanobis, starting the series as the oldest worker present, said that the Institute appeared to him to be like a river which flowed on with contributions made by many workers, some of whom like outgoing branches left after a while, others continuing for years. He then referred to what he described as a special feature of the Institute, namely, the fact that the workers shouldered the burden of many other activities besides the scientific work carried on in the Institute. As an instance of these other activities he mentioned the fact that all the buildings of the Institute had been designed by and constructed under the supervision of the workers. He then expressed the view that it was desirable in the larger interest of the country that the Institute should be run as an independent public institution, so that it could continue to exercise the initiative which had so far marked its activities. Professor Mahalanobis emphasized that statistics should be considered not as an end in itself but as a tool for the advancement of science as well as for the advancement of human welfare—a consideration which in the context of Indian conditions endowed them with very special responsibilities. (See Appendices 2A, 2B, 2C, 2D and 2E for full texts of the speeches of Sri Chintaman D. Deshmukh and Professor Mahalanobis).

8. Professor Mahalanobis was followed by Professor K. B. Madhava, (who had played an important part at the time of inception of the Institute), Sri Jaladhar Sarma (who is the oldest living worker next to Professor Mahalanobis), Dr. C. R. Rao (an old student and at present the Head of the Research & Training School), Sri Mohan Lal Ganguly, Sri Nemai Ghosh and Sri H. K. Chaturvedi (all senior workers of the National Sample Survey Section of the Institute), Sri Samar Mitra (Electronics Laboratory), Dr. Ed. Harper (who is in charge of the Institute's Psychometric Research & Service Unit) and Sri Rabi Sen, an industrialist, who was associated with the Statistical Institute in its early days. The meeting came to an end with the singing of the National Anthem, after which light refreshments were served.

ANNIVERSARY MEETING : 22 DECEMBER

1. The Inaugural Meeting on the 17 December was followed by the Anniversary Meeting on 22 December, which also was very largely attended both by the workers of the Institute and by many guests, among whom was His Excellency Mr. M. A. Menshikov, Ambassador of the U.S.S.R. On the 17th, the proceedings started with the singing of a Vedic hymn, after which congratulatory messages received from important organizations and eminent scientists from all over the world including India were read out (see Appendix 2F). Eminent scientists who were present on the occasion were introduced to the audience by Professor Mahalanobis. At the request of the Chairman some of them spoke briefly.

2. *China :* Mr. Wang Szu-Hua, Deputy Director of the Central Statistical Board and leader of the delegation of four statisticians from the People's Republic of China, conveyed to Indian Statisticians the fraternal greetings of statisticians in his own country. *Japan :* Mr. R. Minobe, Chief of the Statistical Standards Division, Administration Management Agency of Japan, conveyed the good wishes of Japanese statisticians. *Pakistan :* Mr. A. Afzal, Central Statistical Office of Pakistan, conveyed the greetings of the Pakistan Statistical Association. *Switzerland :* Dr. Arthur Linder, Professor of the University of Geneva and the Institute of Technology, Zurich, joined in the expression of good wishes. *Thailand :* Prince M. C. Athiporn P. Ksemasri, Director, Government Statistical Services, National Economic Council of Thailand, brought a message of goodwill from his country. *UK:* Sir Ronald A. Fisher, University of Cambridge, conveyed the congratulations of the International Biometric Society on behalf of its President. Dr. Frank Yates, from the Rothamsted Experimental Station, United Kingdom, conveyed the greetings of his colleagues in England. *USA:* Mr. Morris Hansen, U.S.A. Bureau of Census, conveyed the good wishes sent on the occasion by the Director of Census and the President of the American Statistical Association. *USSR:* Academician N. Bogolyubov brought a message of congratulation from the U.S.S.R. Academy of Sciences. Mr. A. I. Ezhov, Chairman, Scientific Methodological Council and Deputy Director, Central Statistical Board, U.S.S.R., conveyed the cordial felicitations of the following organizations in his country : Central Statistical Board, The Moscow Economic and Statistical Institute and the Economic Department of the State University of Moscow.

3. Other scientists present, who then addressed the meeting, were Mr. F. Khuri from the International Statistical Education Centre, Beirut (Lebanon), Mr. E. Lunenberg of the International Statistical Institute (The Hague), and Professor J. S. Neyman of the University of California (USA).

4. In his speech at the Anniversary Meeting, Professor Mahalanobis emphasized that there was a unity of purpose amidst the great diversity of the Institute's activities. The Institute had carried out theoretical researches not for their own sake but with the object of solving practical problems, which supplied the basic theme.

5. Recalling the persons, who had exercised a deep, though indirect, influence on the affairs of the Institute, Professor Mahalanobis mentioned the eminent scholar, Sir Brajendranath Seal, who had suggested to him as far back as in 1916 that it would be wiser for him to work on statistics instead of continuing his physical researches. It would be difficult, said Professor Mahalanobis, to measure Dr. Seal's influence on the Institute even with the tools of statistics.

6. Professor Mahalanobis then referred to the encouragement received from Rabindranath Tagore at a very critical time during the formative period of the Institute. He mentioned that in 1933 he had received orders for transfer from his post as a professor of Physics in the Presidency College, and that matters had come to a head. He and three or four of his young colleagues (which was the total staff at that time) thought that the best plan would be for him to resign from service and start coaching classes and also to take up agricultural field experiments (using Fisherian design of experiments) on commercial lines and build up the Institute on its own.

7. "It was Rabindranath Tagore", Professor Mahalanobis went on to say, "who gave me complete confidence. I remember, he said, 'Yes, this is the right decision.' He even accompanied me in my search for a plot of land, where experiments could be carried on." He recalled that it was Rabindranath Tagore, from whom he had learnt that in the great past of India, she had opened her doors to guests from all over the world; and in its own humble way, the Institute was trying to follow this example. It would be difficult again, said Professor Mahalanobis, to measure Rabindranath's influence on the Institute.

8. Referring to Sir Ronald Fisher, Professor Mahalanobis said that he during his several visits to the Institute stimulated research in new directions, and in 1939 and again in 1945 he gave his principal support to the Institute's advocacy of the method of sample surveys. Sir Ronald has also always emphasized the importance of the close integration of research, training, and project work, which was the special feature of the Institute.

9. Professor Mahalanobis then mentioned the contribution of the Institute's first President, Sir R. N. Mookerjee, who agreed to accept the office of President at the age of 75, because he felt that India would need more statistics in the future. Sir James Grigg, the Finance Minister in the Government of India in 1935, also helped the Institute to strike roots, when he first sanctioned a research grant of Rs. 5000 for the Institute, which was soon afterwards increased to Rs. 15,000 per year.

10. Professor Mahalanobis next mentioned that in 1943, a considerable part of the work, then being carried on by the Institute, was on the verge of collapse owing to the failure of the Bengal Government to appreciate the value of sample surveys. It was the Honourable H. S. Suhrawardy (at that time the Minister of Civil Supplies and later the Prime Minister of Pakistan), who came to the rescue of the Institute by reviving the grant for the purpose with an imaginative appreciation of the work, which was being done by the Institute.

11. Sri C. D. Deshmukh who became the Institute's President in 1945 at a time of great difficulty had helped in diverse ways, all the details of which, Professor Mahalanobis said, even he himself did not fully know. Finally, he mentioned the support given by Sri Jawaharlal Nehru, Prime Minister, who had an intimate knowledge of the work being done here and without whose support it would not have been possible to carry on at all.

12. Professor Mahalanobis said in conclusion that though he felt a certain amount of anxiety regarding the future, he was confident that there would be nothing to fear, if his colleagues appreciated the inner unity of purpose of the Institute and worked together.

13. In his address from the Chair, Sri C. D. Deshmukh said that in many quarters in this country there was yet no proper recognition of the place that the Indian Statistical Institute occupied in the academic and scientific world. It reminded him of how a philosopher was often not recognized in his own country.

14. Statistics were reputed to be a dry and uninteresting subject, though he could not imagine why it should be so. In essence, it was a science of fact-finding and there was no department of life, in which the formulation of policy or the determination of a course of action did not depend on the ascertainment of relevant facts.

15. Sri Deshmukh observed that ten years ago, there were about 1,000 trained statisticians in India, and the total expenditure on statistics and their application might have been about Rs. 1 crore. At present, the number of workers in the field of Statistics was estimated

by Professor Mahalanobis to be about 4,000 and the total expenditure Rs. 3 crores. It would not be possible to say whether this figure, which indicated the utilization of statistics, was adequate. He expressed the hope that the Indian Statistical Institute would devote a little time to the compilation of comparative statistics regarding the advancement of the science itself and its utilization by public authorities in other countries. "I think we have every reason to hope that, literally, as India's programme of development expands, there will be no end to the vistas and no roof, so to speak, to the Indian Statistical Institute."

16. Paying a warm tribute to the workers, Sri Deshmukh observed, "I have now been associated with this Institute for 15 years and I have seen the numbers grow and I have had many opportunities of watching the selfless spirit of work manifested by the workers of the Institute."

17. In this connexion, Sri Deshmukh mentioned the "Samvadadhvam," in one issue of which workers had proved by some statistical formula that they were paid for doing no work at all. Describing this as an excess of philanthrophy, Sri Deshmuukh said that he was certain that they had worked under very difficult conditions in the past.

18. Sri Deshmukh concluded by saying, "We are now on the threshold of a new age—an age in which there is literally no limit to the useful service that this Institute can render and we are confident that with the understanding and sympathy of all those like you, who are here and those who take an interest in the Institute, the Institute will be able to discharge its duties successfully."

Sir Shri Ram then thanked the guests on behalf of the Institute. The meeting terminated with the singing of the National Anthem. All the guests and workers were then treated to light refreshments.

19. *Lectures and Symposia :* A series of lectures and symposia formed an important feature of the Anniversary Celebrations. These were organized in two groups : the Anniversary Lectures Series held under the auspices of the Indian Statistical Institute and the International Statistical Institute.

Altogether 56 lectures were delivered in the first series starting from 18 December 1956 and ending on 4 January 1957.

The lectures, which were held in the lecture-hall of the Research and Training School, drew crowded audience. Judged by the pointed questions asked, the lectures aroused keen interest.

The International Statistical Seminar Series comprised two symposia on the teaching of statistics, one being devoted to the Training of Statisticians and the other to Training Programme of International Statistical Education Centre, Calcutta. (For details of lectures and symposia see Appendix 18.)

20. *Exhibitions :* An exhibition, depicting with pictures, charts, models etc. and the activities of different departments of the Institute, proved to be a popular feature of the celebrations. There was a section in the exhibition set up by the Central Statistical Organization with charts specially designed to give a vivid idea of the economic conditions of the country.

The Electronics Laboratory organized a separate exhibition, where visitors were shown electronic machines in actual operation.

There was also a photographic exhibition, all the exhibits being the works of members of the Institute staff.

21. *Entertainments :* With a view to making the celebrations popular with, and securing the active participation in them of, all the workers, an elaborate programme of entertainments was arranged and its execution entrusted to the Institute Club. The highlights of the programme were the staging of a Bengali drama and a recital of Indian music, both by the workers, the annual sports meeting and exhibition matches of table-tennis, badminton, volley ball etc. Other entertainments, which proved to be highly popular, were a performance of Rabindranath's *Rakta Karabi* (Red Oleanders) by Bohurupee, a well-known dramatic group in Bengal and a show of the film, *Pather Panchali*.

22. *Excursions :* With a view to giving the guests an impression of the Bengal country-side and of the kind of field-work being carried on by the Institute, a visit was arranged to a village, where the National Sample Survey work was in progress. A steamer trip on the Ganga was also arranged for guests, which proved to be very refreshing after hectic days devoted to lectures and discussions.

23. *News Bulletin :* A news-letter was published daily during the celebrations to keep all visitors posted with the day-to-day programme and other information relating to the celebrations.

24. *Feast :* The celebrations were brought to an end by a feast at Baranagar on 19 January 1957, when about 2,000 workers and guests were served with a sumptuous meal at midday in two relays in the mango-grove. The arrangements were in charge of the Institute Club.

25. *Celebrations in Branches :* The branches of the Institute at Mysore and Giridih celebrated the 25th Anniversary in an appropriate manner in February, 1957.

Appendix 2A

SPEECH DELIVERED BY SRI C. D. DESHMUKH ON 17 DECEMBER 1956 ON THE OCCASION OF THE TWENTY-FIFTH ANNIVERSARY OF THE INDIAN STATISTICAL INSTITUTE

Gentlemen,

I am very happy indeed to have this prvilege of inaugurating the Silver Jubilee Celebrations of the Indian Statistical Institute. I cannot do better than read at the outset a message which has been to us by the Prime Minister, Sri Jawaharlal Nehru. The message said, "The Indian Statistical Institute has gained a reputation for itself not only in India but in every country where statistics are studied. It has grown in recent years and has performed important function in connection with our planning work. Indeed, it is an essential part now of our planning organisation. On the occasion of its Silver Jubilee, I send it all my good wishes."

My first duty on the occasion is to welcome all the guests, particularly those who have come to attend this function or to work in the Institute on the occasion from abroad. They will be introduced a little in detail by Professor Mahalanobis. I would only be content with mentioning that there are about 64 of them from 18 different countries, including 4 children and 12 students for the International School of Statistics. I welcome them all, as they are proof positive of what the Prime Minister has said about the standing and status of the Institute in the international world. I myself have been associated with the Institute for about 16 years out of these 25 and with the Director for about 36 years. Both Professor Mahalanobis and I studied at Cambridge, but he had left, when I joined and I met him for the first time in 1920. Even then, I know that apart from his work proper which was teaching physics, he had started taking interest in matters statistical from the meteorological aspect. I believe he wrote a publication in 1922 and for the next three and a half years he was a meteorologist for the State in addition to his duties as Professor of Physics in the Indian Educational Service. His interest by then had branched out into Anthropology as well as the study of data bearing on rainfall and floods. That interest developed between 1926 and 1931 and it was in 1931 that a meeting of some interested in the subject was called under the chairmanship of Sri P. N. Banerjee, the other members being Sri N. R. Sen and Professor Mahalanobis.

A committee was appointed to direct a plan for the establishment of a learned society or an institute for the advancement of statistics in India. Of that, at that time I was not aware, although by a curious coincidence, between 1926–31, I happened to have carried out certain work of the nature of random sampling for agricultural production as a Settlement Officer in the Central Provinces and Berar which is now a part of Madhya Pradesh. I think it was about the same time that the attention of both Professor Mahalanobis and myself was drawn independently to Havock's work on random sampling. I sensed its potentialities, although I am not much of a mathematician and quite early in my career in statistics I was bowled over Bowley.

In any case, a certain amount of work, thousands of experiments were carried out in random sampling and agriculture by me between 1926–31. I met Professor Mahalanobis again and heard of the development of the society I have referred to in 1939 in Simla, when I happened to be Joint Secretary in what was called the Education, Health and Land Department. It was there that the Professor acquainted me with the difficulties he was encountering in running this institution or Institute or laboratory and the attempts, that he had been making to enlist the sympathy and assistance of the Government of India.

In the meanwhile I think the Institute kept itself going by doing a great deal of work, which the Bengal Government were good enough to entrust to it apart from a certain amount of work carried out for commercial and industrial concerns in the nature of survey and so on and so forth. Attempts to enlist the sympathy of the Government continued from 1939 to 1945 and a great deal of rough weather was experienced in those attempts. As Deputy Governor and Governor of the Reserve Bank, I used to visit Calcutta every cold weather for two or three months, and I took the opportunity of joining the Institute's Governing Body at the invitation of the Council.

In 1945, I was asked to be the President and it was in that capacity that I was able to render some assistance to the Institute in getting a settlement with the Education

Ministry of the Government of India in regard to an annual grant for the Research and Training School, which portion was notionally separated from the rest of the Institute for the purposes so to speak, of attracting a grant. From 1945 to 1950, the Institute carried on with the help and assistance of the Government of India, as also on the basis of certain project work. In between, it was responsible for starting the Quality Control movement in this country in 1948, when Walter Shewhart was invited to visit India in order to acquaint those, who are likely to be interested, in the significance and possibilities of Statistical Quality Control.

Professor Mahalanobis and I came together again in New Delhi in the beginning of 1950, when we discussed the inception of the National Sample Survey, as also the mechanics and the procedure in connexion with the collection of national income statistics. I should say that 1950 marks the end of the struggles of the Institute, because, from that time work came to the Institute, together with recognition in growing volume and the Central Government began to perceive the importance of making greater use of the Institute for the formulation of its policies, not perhaps in the general field of planning so much then, as in the field of policy-making.

The idea that the Institute was capable of helping the Government also in regard to planning came towards the end of the First Five Year Plan period and I believe it was in November 1954 that the Prime Minister inaugurated the work, in which foreign experts participated which led to the publication of the by-now well-known Draft Plan Frame. Now, that in brief is the history of the Institute and the scope of its work. A great deal more will be said on this, not only by some of the speakers, who will follow me but also later during the course of the Celebrations. My purpose has been served by inaugurating these Celebrations and by extending a very cordial welcome to all the visitors especially, as I said, the foreign visitors here.

Appendix 2B

SPEECH DELIVERED BY SRI C. D. DESHMUKH ON 22 DECEMBER 1956 ON THE OCCASION OF THE TWENTY-FIFTH ANNIVERSARY OF THE INDIAN STATISTICAL INSTITUTE

Ladies and gentlemen,

I do not think there is anyone here, who would not have felt an uplifting of his heart on listening to the cordial messages that have been delivered on this occasion from all corners of the earth. There is a saying that when people gather at conferences, sometimes they get conference-minded but I am sure you will realize that this is no ordinary conference. It is a gathering of professional men from the highest levels in their respective countries and as true scientists you may be confident that they will not permit themselves to say anything that they do not sincerely feel. Nevertheless, I recall to myself that a philosopher is not often recognized in his own country and I am aware that in many quarters in this country there is yet no proper recognition of the place that the Indian Statistical Institute occupies in the academic and scientific world.

It is not my intention to ascribe any blame to anyone, because by their very nature statistics are a dry and uninteresting subject. I cannot imagine why it should be so, because in essence it is the science of fact-finding and there is no department of life, in which the formulation of policy or the determination of a course of action does not depend on the ascertainment of relevant facts.

Ten years ago, there might have been about 1000 trained statisticians in this country, and perhaps the total expenditure on statistics and their application might have been of the order of one crore. According to Professor Mahalanobis, the number of statisticians, officially employed, must now be merely 4,000 and the total expenditure Rs. 3 crores. Now, that by itself is somewhat tantalizing, because, unless one has some measure of comparison, one would not know whether the country is dealing adequately with statistics or not. I am not in a position to substantiate the statement that one would find that in other countries a very much larger percentage of the revenues of public authorities are spent on the utilization of the statistical apparatus and organization.

I hope, sometimes the Indian Statistical Institute will devote a little time to this big business of merely compiling comparative statistics in regard to the advancement of the science and its utilization by public authority in other countries.

In any case, I feel we should all be prepared to conceive that India has now put statistics on its map. I should say that a significant event, which pointed to that, was the utilization of the Indian Statistical Institute for the purposes of formulating some kind of a draft plan. I should not like people to run away with the idea that only a statistician can make a plan. It would be very wrong to postulate any such thing.

The statisticians will depend on enumeration of objectives on the authority. The authorities themselves will subject whatever has been prepared statistically to expert examination by economists and the results of the labours of both economists and statisticians will be examined and scrutinized in the offices of the Government from the administrative and, later, from the political point of view and then only will some kind of proposal emerge, which can be called a plan or a programme of action. I am very happy to be able to inform you that in the processes that led to the formulation of the Second Five Year Plan, there was an almost perfect cooperation between all our relevant internal agencies as well as collaboration by a very large number of devoted and interested scientists from abroad. In these circumstances, I think, we have every reason to hope that literally as India's programme of development expands, there will be no end to the vista, as I said the other day and no roof, so to speak, to the Indian Statistical Institute.

Professor Mahalanobis is not an easy man to work with. He is so interested, so absorbed in his work that he cannot resist any line of advance that promises results for the advancement of the purpose in view and he is not a believer in statistics being a kind of fundamental science. He also does not take a very narrow view of statistics as an applied science. All the time he has one eye on the science and another eye on the application of the science to the solution of the big current problems of the day and he casts his net very widely. Every time he comes back from a foreign tour, I dread to hear of some plan of development. Sometimes it is Quality Control; sometimes it is Production Engineering; sometimes it is Perspective Planning and then it is the Electronics Computer and, since progress in the world also is continuing very fast, the number of ideas he brings back is larger every year, but I think that as it should be. After all here is a challenge, which workers and colleagues and the practitioners and the teachers of the science can take up. The challenge is one that is easy to understand. It is a fascinating adventure, in the words of the Prime Minister, this work of developing the biggest democracy in the world or one of the biggest democracies in the world and in any case advancing the interest and lifting the living standard of the State—is it one man in five or one man in six—I do not again know what is statistically correct. At any rate that is a large enough human responsibility and, if we have to take a view, I should say that I would rather err with Professor Mahalanobis than be right with somebody else arguing for conservatism and orthodoxy in a matter of this kind. And, therefore, as somebody said the other day, 25 years is at least one third of a man's life and in the life of an institution one does not know what it is; it might be one of the smallest of fractions. Therefore, there is no limit to the assistance that this Institute can give to the authorities of our country. That has been realized by the Central Government. They themselves have offered to the State Government to share the cost of expanding their statistical organization, so that all the information, which is relevant to measuring the execution of the Plan, should be collected and garnered. It should be in the hands of those policy-makers and the supervisors and those finally, who are directly responsible for the execution of individual projects. Indeed, the more plentiful is the systematic collection of facts, easier it becomes to delegate work. Same is the matter in the industrial world. If there is job analysis and in many countries almost every job is analysed, you can find out what the norms for work are and once you know what the norms of work are, you can find out who is doing his best and who is an idler and after all, in scientific jargon, that is the essence of progress in any country, that is to say first to enable people to work and secondly to ensure that, when proper conditions have been established, people do discharge their work. Therefore, as I said, the Central Government has realized this, many State Governments have accepted the offer made by the Central Government and we hope that in course of time the Institute will be able not only to carry out projects for Government but also to help in the broadening and widening of statistical organizations at different physical levels, so that the business of implementing the Plan is watched and measured in an ordinary fashion. As a result of this understanding many of the old conflicts are gradually disappearing. Sometimes, conflict occurs at the scientific level, as for instance, whether the samples taken for random sampling are too small or too large but all these things the scientists learn to leave behind them and, once the basic purposes are realized, then in spite of diversity of work, as Professor Mahalanobis has said, the unity of purpose triumphs and always remains to the fore and those are the conditions in which all statisticians which mean all workers—we, here, have no distinction of posts and we have no hierarchy except for the discharge of functions—can work with the same spirit.

I should like to take this opportunity of paying my tribute again. I did it the other day, the whole of the last meeting of the 17th was a tribute to the workers. I should like to pay a tribute to them again. I have now been associated with this Institute for 15 years and I have seen the numbers grow and I have had many opportunities of watching the selfless spirit of work manifested by the workers of the Institute. Indeed, they have gone to the length of proving in their second issue of 'Samvadadhvam' that they are paid for doing no work at all; they have given some statistical formula, by which they have proved somehow that there are so many holidays and there are so many close days and they have statistically proved that they are paid for doing no work. But I regard that as an excess of philanthropy. Leaving that apart, I do realize they have worked under very great difficulties and, what is more, sometimes things have appeared very uncertain to them and their future has appeared very gloomy and dark.

I have reasons to hope that those times are ever and that we are now on the threshold of a new age—an age in which there is literally no limit to the useful service that this Institute can render and we are confident that with the understanding and sympathy of all those like you, who are here and who take an interest in the Institute, the Institute will be able to discharge its duties successfully.

Appendix 2C

CONCLUDING SPEECH BY SRI C. D. DESHMUKH ON 17 DECEMBER 1956 ON THE OCCASION OF THE TWENTY-FIFTH ANNIVERSARY OF THE INDIAN STATISTICAL INSTITUTE

Ladies and Gentlemen,

This brings the proceedings of this evening to a close and I shall only make a few concluding remarks.

Reference has been made in the course of the speeches to the rapidity of the growth of the Institute in recent years. I do not know if you are aware of the latest example of this rate of growth, which is the construction of two Guest Houses within the space of three weeks, thanks to the energy of the worker in charge, Sri Subodh Das Gupta and the personal supervision of Mrs. Mahalanobis. Many people, who visited the Institute about a month ago and are visiting it again, are astonished at the appearance, as if by the rubbing of Aladdin's lamp of these two Guest Houses.

I hope you have not been bored by the many speeches that you have heard ; but were arranged with this object that you should hear of the personal feelings and experiences of the different strata, if I might use the word, of the workers belonging to different periods in the history of this Institute. I am sure, you all have been glad to notice the spirit of camaraderie and the faith and self-confidence which was illustrated by the observation made by the workers.

The history of the Institute falls into several periods: the first one from 1920–31 —the period of gestation; the next from 1931–39 was the period of the struggle of infancy; 1939–45 consisted of the disappointments and frustrations of adolescence; 1945–50 were characterized by the hopes of youth; and from 1950, we may say, the Institute has been looking forward to the wider horizons of manhood. With the recognition in principle by Government of the all-India and indeed the international importance of the Institute we are standing on the threshold of a period of endless vistas. I am quite sure that at the rate at which the Institute is growing, it will be able to discharge all the functions that it may take on itself for carrying out its double responsibility or fulfilling its double purpose, to which reference was made by Professor Mahalanobis. I should like to underline what he has said in regard to relations with Government and the importance of autonomy. What he said would have appealed to me in any case, because in my present capacity as Chairman of the University Grants Commission I have some understanding of the great importance, almost indispensibility, of allowing academic bodies a free hand in regard to the management of their academic work and generally in regard to their day-to-day administration. This is a factor of the utmost importance. If any kind of administrative constraint were to be placed on the Institute, I am quite certain that the results will be very disappointing and self-defeating, in that the Institute will not be able to come up to expectations. One feature, of which the Institute might be justly proud, is the extent to which it has enlisted international cooperation. I am not aware of any other learned body or any other academic institution, which has attracted the help of outside experts to such a large extent as this Institute. There is a continuous give-and-take exchange of visits with consequent enrichment of the work turned out by the Institute.

The Institute is in the peculiarly useful or valuable position of being able to assist Government in ways, in which perhaps no other form of organization can assist it. There is, for instance, the question of study in matters connected with Government's financial and planning policies and yet matters which do not fall either within the category of research or within the category of day-to-day administration. The latter can be carried out by Government's own economic adviser, statistical adviser and others. Research can be carried out both in the Research and Training School of the Institute as well as in the statistical departments of the universities.

But there is a field in-between, where full thought has to be given to the problems of policy, not so much in the sense of day-to-day decisions as in the sense of defining the conditions, under which policies might be implemented and that kind of work can best be carried out by an Institute of this character. I do hope that in the future in addition to work that is carried out in connexion with planning, as for instance, perspective planning, the Institute will be utilized by Government for the purpose, that I have just indicated. I hope that in the course of these Celebrations those, who are interested, will have the opportunity of listening to the observations of most of our other visitors not only from the point of view of friends and sympathisers of the Institute but also experts, who have brought to us their own special fund of experience.

Thank you.

Appendix 2D

SPEECH DELIVERED BY PROFESSOR P. C. MAHALANOBIS ON 17 DECEMBER 1956 ON THE OCCASION OF THE TWENTY-FIFTH ANNIVERSARY OF THE INDIAN STATISTICAL INSTITUTE

I will start the series of talks about old days, in which my colleagues will join. My mind goes back to the day, on which we started our work; and I have been thinking what I shall tell you today, after 25 years. I shall speak particularly to my fellow-workers.

One thing which is constantly in my mind is that the Institute is somewhat like a river. Men come and work; some continue for a long time; some stay a short time and go away. But this common endeavour of different individuals flows on as the activity of the Institute.

I am keenly conscious that many individuals with many different interests have built up our Institute. Circumstances have forced us to make the Institute highly self-contained in nearly every thing. This is one special feature. It is not that we are doing only scientific work; the workers of the Institute shoulder the burden of many other things. I am very proud that in this Institute our buildings have been all designed and constructed under our own supervision. We run many things, and this is why I feel very deeply that even those workers, who are not directly concerned with statistics, also belong to this Institute in a very special sense. This is one thought, which I should like to convey to you.

Statistics has a very wide field. It has many aspects and in this Institute it has given us the opportunity for cooperative work in a very broad sense, in the building up literally of this Institute, as witnessed by these buildings, these grounds and, of course, the scientific work.

I am also conscious of the Institute's very rapid growth, which is a matter of congratulation in one sense, but at the same time I also see the dangers of too rapid a growth. There is danger, when we grow too big or too rapidly; our work tends to become rather impersonal and we begin to lose human values which is a great loss—this also has been uppermost in my mind today.

I remember the early days, when we started with one part-time worker. Then we had a small group of four or five. For the first three or four years, I think, the total number did not exceed 10, or may be 12. We had a typist, Jaladhar, who is now our oldest worker and who will speak of the early days very shortly. We also had considerable help from outside. Professor K. B. Madhava, for example, played an important part in shaping the policy by insisting that the name should be "Institute" and that we should not be merely a talking society but we should take up research and even training. The idea was that the Institute did not merely mean coming together to read papers but something more, something that would cover a wide field of activities. In this way, we began to have many workers, who perhaps did not publish important papers, but who in a way contributed much that was of value to this Institute. Today I remember Sudhir Banerjee who died in 1952. In our early days the Institute could scarcely have survived without his very wise management of financial and business matters.

In many different ways we have grown up and our responsibilities have increased. Twentyfour years are at our back. In looking forward I feel more and more the need of strengthening the spirit of cooperation and developing some unity of purpose together with, of course, a great deal of diversity in our individual way of doing and looking at things.

I should like to touch very briefly on one aspect of the Institute's relations with Government, which has been under consideration for twenty years. Our President has referred to this. We have continually pressed on Government the need of giving the Institute full freedom and autonomy in internal management. We are a non-government but public institution; and, so far, we have tried to discharge our public responsibilities not as a Government department but as an independent institution. I am myself convinced that this is desirable not so much in the interest of the Institute itself but of the country as a whole. In fact, converting the Institute into a Government department would relieve us of much worry, because things will go on somehow but there would be the danger of losing the initiative. Therefore, I think it is important in the best interests of the country that we should continue to function as an independent public institution. I am glad to say this has been accepted in principle, and we are hoping that fairly soon it would be recognized through parliamentary action.

But it is the other side of this responsibility, which I should like to impress on my colleagues. If we claim a certain amount of freedom of action, then we must also shoulder a certain amount of responsibility, that is, we must see that our work from within is maintained at a high level of efficiency, and secondly, that we retain our initiative—the initiative to direct our efforts, our united endeavour towards solving either scientific problems or problems of national development, which will be continually facing us in India.

I do believe that our Institute has a special responsibility, because we happen to be located in India. There are many problems which face us in India, which perhaps do not face countries which are more advanced; and, therefore, if we consider statistics not as an end in itself but as a tool for the advancement of science as well as for the advancement of human welfare, then there must be special responsibilities determined by our special needs in India.

I have very much in mind today this aspect of statistics and, therefore, we must never lose the initiative. We must remember that in an under-developed country statistics must have a clearly defined purpose, one aspect of which is scientific advance and the other, human welfare and national development.

This is my appeal to all my fellow-workers at this moment, when we are getting over the first quarter of a century of our activities. It was on 17 December 1931 that we had the meeting, at which it was decided to establish this institution. Let us look forward and shoulder our responsibility to do the best we can to improve our own efficiency of work, and to maintain our initiative in solving problems of national development in future.

Appendix 2E

SPEECH DELIVERED BY PROFESSOR P. C. MAHALANOBIS ON 22 DECEMBER 1956 ON THE OCCASION OF THE TWENTY-FIFTH ANNIVERSARY OF THE INDIAN STATISTICAL INSTITUTE

Respected President and friends,

I can claim to be the oldest worker of the Institute. As I look back to the last 25 years, or even longer, because we did have a Statistical Laboratory with a small group of workers before the Institute was founded, several thoughts come to my mind.

Firstly, I believe, I can see some unity of purpose in the great diversity of our activities. I may perhaps say that whatever work we did in this Institute have been mainly directed to solving problems at the practical level; but we have found that solving a practical problem continually called for a good deal of theoretical researches. Therefore, we took up theoretical researches not so much for its own sake but to meet real needs of solving practical problems.

There is unity in this approach. There is no conflict, I think, between scientific research and solving practical problems. Perhaps, we may look upon scientific curiosity itself as an emergence in the human mind of an urge to know more of nature and, through such knowledge, advance the welfare of the human species. If one looks at it in this way, there is no conflict between scientific research and solving practical problems.

Another way of appreciating the unity of the work of the Institute, I think, stems from the deeper physical, social and economic needs of an under-developed country like India.

Some of the early problems, on which we worked, were concerned with rainfall and floods and what remedial measures could be taken. Then came enquiries into rural indebtedness, handloom weavers, agricultural crop surveys, and later on, the collection of social and economic information for this vast country, in which we have ten or perhaps twenty or thirty million persons unemployed or under-employed according to the definition of unemployment we use. These enquiries all stem from the needs of an under-developed country; and still more recently from the needs of planning.

This is how gradually the Institute expanded its activities. Perhaps, the unity as well as the diversity may be appreciated in the vivid contrast between our modern workshop or our unit for electronic computation, which are located in buildings which adjoin *Kalyanasri*, the unit for the study of cottage industries like the *Ambar Charkha* (spinning wheel) and handloom weaving. Both types of activity have the common aim of using science for the promotion of national development. The *charkha* and the electronic computer working side by side in our Institute make vivid the unity of purpose in the midst of sharp contrasts.

My mind goes back to the problems, which we have taken up for study. All are fundamentally problems of practical importance, but we worked only on such problems, for which we had men to take up the challenge. In the Institute we have never thought in terms of posts to do this or that. We have always thought of men, who could do creative work; and it is round men that the work of the Institute has grown up. I think this is an important feature of the Institute. It is something like a stream of activities of individuals. The Institute started work with only one part-time computer 25 years ago. The Institute has grown gradually like a river gathering tributaries, by opening out branches and sending out men to other agencies. This has given us an inner sense of direction, not in a mechanical way, but through a kind of unity of purpose which is living. I should stress this as an important feature, as I look back. The history of the Institute then is really the response of human beings to the need of solving problems of real life. This, I should say, is our unity of purpose.

I, therefore, turn my mind to the men, with whom I have had the privilege of working. Here also there is a great contrast. If we look today at the flood-lighted building of the Institute we see only the facade. The foundation cannot be seen; neither the roof at the top. As a matter of fact, the foundation and the roof must not be seen. Their fulfilment is in not being visible. In the same way, I am aware, as the oldest worker of the Institute, of the foundation, which has been built by humble men, whose name nobody knows. I recall the many with whom I have worked. Now the Institute has grown so big that I am sorry I have no longer that privilege. But it is the dull, day-to-day routine work of many humble men, which has built the Institute. It is wise to remember this today.

Then there are others who exercised a deep influence, but not in a direct way. They are not visible in the list of papers which we have published. I shall recall some of them to convey to you what I have in mind today. Brajendra Nath Seal, the great Scholar with whom I started doing some statistical work, had an imagination which is still fresh in my mind as something wonderful. He told me that India would need modern statistics, and advised me to work in this subject instead of doing physical research. This was in 1916, 40 years ago. It was his advice and encouragement which set my course. His influence on the Institute is difficult to measure with the tools of statistics.

I should like to name another great personality, Rabindra Nath Tagore, who gave us great encouragement, and who used to visit the Institute frequently. Very soon after the Institute had been established, I recall, we were doing much advisory work in the design of agricultural field experiments. I was professor of Physics in the Presidency College, which is a Government institution in Calcutta; and I learnt that Government intended to promote me and send me out of Calcutta as the principal of a college. This would mean the end of the Institute and statistical work. At that time, we were a small group; Subhendu Bose and Sudhir Banerjee and 3 or 4 others. We were not rich and we were in difficulties. We thought over the matter carefully and felt that the best thing would be, if I resigned from Government service and then we could start coaching classes in the evening and continue our statistical studies. It was Rabindra Nath Tagore who gave me encouragement. I remember, he said, "Yes, this is the right decision." He even went with me to find an old garden house (very near this place), where we could do our agricultural experiments. Again, I do not know how we shall measure his influence. Not only this. It was Rabindra Nath Tagore, who made me realise that in the great days of India she had opened her doors to the whole world and had welcomed guests from abroad. This is something for which we are truly grateful in this Institute, because we also have had the great privilege of having many friends from abroad. I do not know how to measure the help received from them.

I may turn to Sir Ronald Fisher. His influence on this Institute is indeed difficult to measure. It has been not only in stimulating research but also in the support he gave to our advocacy of the sample survey or again to the need and the importance of a close integration of research, training and project work. He gave us support at a very critical stage of our work in 1938; and his influence is something which is deep although it is not visible on the facade.

There have been also some men at the top who have helped the Institute in ways which people do not know. I may mention our first President, Sir Rajendra Nath Mookerjee. When we wanted to start the Institute, I did not know that an attempt had been made earlier to establish a statistical society, which came to nothing because of the lack of confidence and understanding between the Indian and British businessmen. It was only because Sir R. N. Mookerjee agreed to become our President that it was possible to start this Institute on a broad basis and which soon made it possible for the Institute to secure a public status. He was old, 75 or 76; and was retiring from all offices. But he made one exception and became our President. One day he told me, "Well frankly, I do not quite understand all you are saying; but somehow I have a feeling that India will need more and more statistics in future. This is why I am helping you." It was his imaginative understanding which gave us a start.

I also recall Sir James Grigg, then Finance Member in the Government of India. He came to see the Institute in 1934. We had a grant of Rs. 2,500 a year at that time. All on a sudden he asked me when he was getting away in the car, "Do you want a research grant ?" I said, "Yes". He said, "How much ?" I could not think of anything more than Rs. 5,000. He immediately sanctioned Rs. 5,000. A little later, even before we started getting the new grant, he increased it to Rs. 15,000 a year, which gave us the opportunity to start a small nucleus of whole-time workers. This gave us the opportunity of striking roots.

I shall mention another name. At a time of great difficulties in Bengal, in 1943, it seemed that the work of the Institute would collapse. We had built up the sample survey of crops after six years of hard work. In 1943, there was a new Agriculture Minister of the Government of Bengal; he did not have any faith in sample surveys; he could not simply understand how by taking observations on one out of 200 fields one could get an estimate of the jute or the rice crop. So he stopped the sample survey. At this time, the Minister of Civil Supplies, the Honourable Mr. H. S. Suhrawardy, later the Prime Minister of Pakistan, came forward to the rescue. He was in need of reliable information on the supply of rice. He showed an imaginative appreciation of the work of the Institute; and he sanctioned the grant. In this way the Institute received help from many unexpected quarters.

I should like also to say a few words about our President whom I have not mentioned before. He became associated with this Institute, I think, in 1940 or 1941 as Vice-President. He became our President at the time of a grave crisis in our affairs in 1945, when we were under great pressure from the Government of India. Since then he has been helping us quietly in many ways, some of which even I never know. This is, I think, all I need say of him.

Finally, I should like to mention Jawaharlal Nehru, our Prime Minister. We have had the privilege of having him as a frequent visitor in this Institute. He has stayed here; and he knows the work of the Institute intimately. He has given us strong support, without which we could not have gone on with our work. Two years ago he came here; and in the Institute he inaugurated the studies on planning for national development. He is one of the men at the top whose name will not appear in our formal history.

When I look to the future two thoughts come to my mind. When we have had individuals to meet the challenge of a new problem we had nothing to fear. Secondly, in this Institute, our best work was done when we could somehow, perhaps not quite consciously, realize the unity of purpose, and could cooperate among ourselves. On the other hand, we have passed through dangerous days of stress and strain, especially after rapid expansion when disruptive tendencies arose. One such crisis had come 8 or 9 years ago. Recently we have been again expanding at a rather alarming rate. So on this day, as I look to the future my mind is not completely at ease—I have some misgivings. The future depends on us—my fellow-workers. If we can realize the unity of purpose and work together, then, I think, there is nothing to fear. This is the thought which is uppermost in my mind today.

One final word. We are fortunate today in having with us colleagues and friends from all over the world, East and West, North and South. I recall an old text of my country from the 'Atharvaveda.'

Yatra visvam bhavati ekanidam,

which described a civilised country as a place "where the whole world meets and finds its one nest." This text gives me courage and hope; and I salute all our guests and friends from abroad.

Appendix 2F

LECTURES, SEMINARS AND SYMPOSIA DURING THE TWENTY-FIFTH ANNIVERSARY CELEBRATIONS : DECEMBER 1956 TO JANUARY 1957

18 December 1956 :

PROFESSOR T. KITAGAWA (*Japan*) : The logical aspects of successive processes of statistical inference.

SIR RONALD A. FISHER (*UK*) : Statistics in scientific work.

19 December 1956 :

DR. Q. M. HUSSAIN (*Pakistan*) : Recent contributions to design of experiments.

DR. U. S. NAIR : Study of maternity rates.

PROFESSOR T. KITAGAWA (*Japan*) : Application of design of experiments to Japanese engineering.

SIR RONALD A. FISHER (*UK*) : Probability in the sciences.

20 December 1956 :

PROFESSOR V. M. DANDEKAR : Fisherian inference.

SRI N. CHAKRAVARTI : Organization of a State Statistical Bureau.

DR. M. ZIAUDDIN (*Pakistan*) : On symmetric function statistics.

PROFESSOR S. K. TAI (*China*) : 1953 Population census of China.

DR. F. YATES (*UK*) : Statistical aspects of agricultural planning.

PROFESSOR S. H. WANG (*China*) : On the coordination between the industrialization and agricultural cooperation.

SIR RONALD A. FISHER (*UK*) : Probability in the sciences.

PROFESSOR J. S. NEYMAN (*USA*) : Statistics, the servant of all sciences.

21 December 1956 :

DR. B. D. TIKKIWAL : Successive sampling and its role in sample surveys.

ACADEMICIAN N. N. BOGOLYUBOV (*USSR*) : Some remarks about the problem of statistical equilibrium.

DR. NATHAN KEYFITZ (*Canada*) : Technical assistance and economic planning.

PROFESSOR A. LINDER (*Switzerland*) : Principles of experimental designs.

MR. E. DA COSTA : Regional distribution of Indian National Income.

DR. V. G. PANSE : Development of statistics in relation to agricultural research, especially plant breeding in India.

22 December 1956 :

MR. N. T. MATHEW : Non-response errors.

DR. A. K. GUPTA : Statistical methods in the work of standardisation.

PROFESSOR J. NEYMAN (*USA*) : Inductive behaviour as the basic concept of philosophy of sciences.

PROFESSOR D. R. GADGIL : Some aspects of economic planning in India.

DR. A. I. EZHOV (*USSR*) : Organisation of the State Statistics in the USSR.

MR. M. HANSEN (*USA*) : Application of electronic computers to data processing.

DR. N. S. R. SASTRY : Rural Credit.

24 December 1956 :

DR. B. RAMAMURTI : Statistical system in India.

PROFESSOR K. N. RAJ : The working of the money income multiplier.

PROFESSOR K. IYENGAR : Some present-day problems of low income urban workers in India— a realistic approach.

PROFESSOR D. G. KARVE : Community projects and National planning.

Chairman : SIR RONALD A. FISHER (*UK*) : Symposium on Teaching of Statistics and Training of Statisticians. *Participants :* DR. P. B. PATNAIK, DR. Q. M. HUSSAIN, DR. M. ZIAUDDIN (*Pakistan*), DR. U. S. NAIR, SRI V. M. DANDEKAR, SRI D. Y. Lele, PROFESSOR V. D. THAWANI, DR. Q. K. GAYEN, PROFESSOR P. C. MAHALANOBIS, SRI N. T. MATHEW, DR. B. RAMAMURTI, DR. B. D. TIKKIWAL, PROFESSOR T. KITAGAWA (*Japan*).

26 December 1956 :

DR. A. K. GAYEN : Rank analogue of the standard tests of significance. *Chairman* SRI A. C. GUHA : Some techniques of decentralised control. *Speaker :* PROFESSOR J. P. MILLER (*USA*).

Chairman : PROFESSOR A. LINDER (*Switzerland*) : Symposium on the teaching of Statistics—Training of the International Statistical Education Centre, Calcutta. *Participants :* MR. R. MINOBE (*Japan*), MR. FAIZ EL KHURI (*Beirut*), PRINCE ATHIPORN KSEMASRI (*Thailand*), DR. B. RAMAMURTI, DR. NATHAN KEYFITZ (*Ceylon*), DR. P. B. PATNAIK, DR. M. ZIAUDDIN, DR. Q. M. HUSSAIN, MR. A. AFZAL (*Pakistan*), DR. A. MATTHAI, PROFESSOR P. C. MAHALANOBIS, PROFESSOR A. LINDER (*Switzerland*), SRI S. SENGUPTA, SRI S. K. BANERJEE.

27 December 1956 :

DR. C. CHANDRASEKHARAN : Demographic surveys in underdeveloped countries.

MR. D. J. DESMOND (*UK*) : Operational research ideas in sampling plans.

MR. A. L. RAICH (*USA*) : Some Industrial problems for the mathematical statisticians.

28 December 1956 :

MR. R. P. BHARGAVA : Forest sample surveys.

PROFESSOR A. LINDER (*Switzerland*) : Statistical aspects of Salks' vaccine.

DR. P. K. BOSE : Some statistical problems associated with scholastic tests.

DR. M. R. SAHANI : Man and his ancestry through geological age (with lantern slides).

29 December 1956 :

PROFESSOR M. C. CHAKRAVARTI : Teaching of designs through matrix methods.

DR. A. R. KAMAT : On some statistics based on the variate differences.

DR. N. V. A. NARASINHAN : On the structure of the Indian economy.

31 December 1956 :

DR. V. S. HUZURBAZAR : Statistical inference.

PROFESSOR B. R. SHENOY : Foreign aid and Indian economic development.

Dr. D. N. Majumdar : The mystery of a Himalayan lake— Rupkund expedition (with lantern slides).

2 January 1957 :

Dr. J. N. Mukherjee : Scientific approach to planning.

Dr. Gyanchand : Premises of the Marxian view on population— a critique.

Professor K. P. Chattopadhyaya : National Planning and culture change.

Professor J. Maclean (*UK*) : Statistics and arithmetic in primary schools.

3 January 1957 :

Dr. S. K. Banerjee : On the applications of statistical methods to medium range and extended range weather forecasting.

Dr. V. K. R. V. Rao : Some problems of long range planning.

4 January 1957 :

Dr. U. P. Basu : On the planning of chemical industry.

Professor P. Roy : The pattern of scientific and industrial research.

5 January 1957 :

Dr. N. R. Sen : The statistical theory of isotropic turbulence.

Appendix 2G : Congratulatory Messages

Mr. A. M. Nesmeyanov, *President of the Academy of Sciences, USSR* and Mr. A. B. Topchiev, *Chief Scientific Secretary of the Presidium of the Academy of Sciences, USSR :*

The Academy of Sciences of the USSR on the day of the 25th Anniversary of the Institute, of which you are the head, conveys its most sincere congratulations to you and to all the workers.

By its outstanding scientific activities and by its important role in the social life of the country, the Indian Statistical Institute has merited high respect far away from the boundaries of India. In this famous epoch of the revival of your country, the Indian Statistical Institute has made an important contribution to the scientific planning of the development of Indian economy and national welfare.

The broad scientific ties of the Institute with scientists of different countries and the scientists of the USSR in particular have become glorious traditions. Such scientific ties and personal contacts among scientists are of great value and we hope they will develop and strengthen in the future.

The Academy of Sciences of the USSR wishes new success to the Indian Statistical Institute which will enrich the treasures of national and world culture.

Mr. Ryiabushkin, *Director of Statistics, Institute of Economics, Academy of Sciences, USSR :*

The Statistical Section of the Institute of Economics of the Academy of Sciences of USSR sends you and your colleagues warm greetings for the Jubilee of the Institute and wishes you and your colleagues new creative successes in the field of Statistical theory and its applications.

Mr V. Ovsienko, *Rector, Moscow Economics and Statistics Institute :*

In connexion with the Twenty-fifth Anniversary of the Indian Statistical Institute, Moscow Economics and Statistics Institute send its best wishes for development of Statistical work towards higher success and better life of the Indian people.

Professor M. I. Rubinstein and Madame Rubinstein, *Institute of Economics, Academy of Sciences, USSR :*

Congratulate significant date 25 Anniversary Indian Statistical Institute wish further success your research relating Indian Economic Development best greetings to Rani and all Friends.

Mr. Ezhov, *Chairman of the Council of Methodology and Deputy Director of the Central Statistical Bureau, USSR :*

On behalf of the Council of Scientific Methodology under the Central Statistical Bureau USSR and personally send greetings and congratulate you and your colleagues for the Jubilee of the 25th Anniversary of the Indian Statistical Institute. The many-sided activities of the Institute have already acquired world-wide recognition and appreciation.

The great work conducted by the Institute under your guidance particularly in the sphere of application of sampling methods is studied with interest in many countries. Great significance lies in the fruitful training of highly qualified statisticians as well in the sphere of exchange of experience and strengthening of scientific collaboration among the statisticians of many lands including the collaboration among statisticians of India and those of the USSR. The Soviet Statisticians are in deep sympathy with the activities of the Institute directed to influence the development of the national economy and raising of the material and cultural standards of the people of India. The Soviet Statisticians are glad at heart at the successes achieved by the Institute in this task. The Council of Scientific Methodology sends heartfelt greetings to your colleagues and you and wish you greater success in improving statistical methodology and practical application of statistics towards the welfare of the great Indian nation for peace and progress of the whole world.

Mr. D. Degtyar, *Phnom Penh, USSR :*

I am very glad to have an opportunity to convey my congratulations to you and through you to all the staff of the Statistical Institute with the 25th Anniversary of the Institute.

I have had the pleasure to work in your Institute and to observe with the great satisfaction the great work which is being conducted by the Statistical Institute in the field of Planning of New India's economic development as well as in the strengthening of the contacts between the scientists of various countries and especially between the scientists of India and the Soviet Union.

Staying at the present time in the Kingdom of Cambodia I wish with all my heart successes in the further fruitful work to the Statistical Institute on the day of its anniversary, in the work directed to the development of your country on the way of progress and strengthening of economic independence.

ACADEMICIAN K. OSTROVITYANOV, *Vice-President, USSR Academy of Sciences :*

Congratulate you and staff of Indian Statistical Institute on occasion remarkable date of 25th Anniversary of the Institute. Under your guidance Indian Statistical Institute has become one of world's largest research centres in the field of statistics. For the last few years the Institute has effected extensive and important research into economics and applications in prospective planning of peoples' economy of India. The Institute has become place of current meetings, fruitful discussions and cooperative work of Statisticians and Economists of different countries. The works done serve vivid example of possibilities for international cooperation in social and economic sciences. I wish Indian Statistical Institute further successes in the fruitful research and achievements of Indian Scientists for further progress of economy of your country and rise of living standard of Indian People.

PROFESSOR V. STAROVSKY, *Director of the Central Statistical Bureau, USSR :*

On behalf of the Central Statistical Bureau, USSR and of all the workers of the organs under the Statisticians' Union (Government) of USSR and on my own part, I hereby send greetings to you and your colleagues and the statistical workers of India on the occasion of the 25th Anniversary of the Statistical Institute for the excellent work in both the spheres of theory and practice of statistics conducted under your guidance by the Indian Statistical Institute and particularly for the famous National Sample Survey undertaken for the welfare of the people of India and for the training of highly qualified statisticians. Such extensive activity of the Institute is well-known among wide circles of Soviet Statisticians exciting their admiration and sympathy for the achievements of the Institute. Wish you greater successes in your fruitful scientific and practical work directed to the welfare of the people of the Indian Republic, and also heartily wish for further strengthening of the bonds of friendship and collaboration among the statisticians of India and USSR for peace and progress of the world.

MR. LAPTEV, *Director of the Institute of Economics, Academy of Sciences, USSR :*

The Director of the Institute of Economy of the Academy of Sciences USSR sends warm greetings to you and your colleagues on the occasion of the Jubilee of the Institute. The Directorate wishes all the members of the Institute and you personally further creative successes in work.

PROFESSOR I. UDALTZOV, *Dean, Economic Faculty, Moscow State University :*

The Economic Faculty of the Moscow State University takes the pleasure of sending you and the Institute under your guidance heartiest congratulations on the 25th anniversary of the Institute.

YUGOSLAV STATISTICAL SOCIETY :

Please accept best wishes, occasion 25th Anniversary Indian Statistical Institute.

MR. YOU POH SENG, *University of Malaya :*

I am sorry to miss this further opportunity to revisit the Institute, where I had had some very pleasant memories from my previous visits. I do, however, offer my sincerest congratulations to the Institute and to all connected with it, and to you especially, on this occasion of its 25th Anniversary, and I wish every success to the celebrations and to the symposium, which is scheduled to follow this celebration.

Mr W. R. Leonard, *Director, Statistical Office, United Nations, New York :*

I should have very much welcomed the opportunity to be present. The Institute has had a very distinguished history and in recent years is playing as ever increasing role in Indian affairs as well as in International affairs. I wish the Institute continue an expanding success.

Mr. R. F. Fowler, *Ministry of Labour and National Service, London :*

My warmest congratulations on the 25th Anniversary of the Institute, and my best wishes for the next 25 years.

Dr. J. D. N. De Fremery, *Secretary, Verenging Voor Statistiek, The Hague :*

Wishing a successful meeting at your Anniversary Celebration and the best result for your work in the future.

Mr. Herbert Marshall, *Dominion Statistician, Dominion Bureau of Statistics, Ottawa, Canada :*

We shall be thinking of your meeting in December. Wishing we could have been there and hoping that Anniversary will be eminently successful, as, I am sure, it will.

Professor Charles Bettelheim, *Professor of Economics, Ecloe Pratique des Hautes Etudes, Sorbonne, France :*

On the occasion of the twenty-fifth Anniversary of the Indian Statistical Institute I am sending to you my best wishes for its future development and for the workers of the Institute.

Having been associated with the activity of the Indian Statistical Institute during nearly three years, I know how much valuable work is being done there in the most various fields and not only in the field of statistics. I know what important part the Indian Statistical Institute is playing also in the field of international cultural and intellectual relations.

I wish to the Indian Statistical Institute, with its already long past behind it, a still more fruitful future.

Professor Darmois, *University of Paris :*

I want to state here that my spirit will be present with you on the 17th December 1956 to felicitate you and rejoice over the magnificient creation which you have achieved, not only of an Institute flourishing on its results, but also of an excellent school of statisticians.

MR. C. J. MARTIN, *The East African Statistical Department, Kenya :*

May I send my sincere best wishes for your Anniversary Celebrations and congratulate you on the great achievements, which the Indian Statistical Institute have made in the 25 years of activity ?

SRI G. L. NANDA, *Minister for Irrigation and Power, Government of India :*

Kindly convey to the workers of the Institute my deep appreciation of the work done by them; especially that connected with the formulation of the Second Five Year Plan.

SRI AJIT PRASAD JAIN, *Minister for Food and Agriculture, Government of India :*

I am glad to note that the 25th Anniversary of the Indian Statistical Institute is going to be celebrated this month. Your Institute has made a valuable contribution to the statistical work in India.

I wish your celebrations success.

Messages of greetings and good wishes were also received from Mr. M. H. Campion, Director of the Central Statistical Office, London; Dr. D. Ensminger, Ford Foundation, New Delhi; Mrs. Karin Kock, Head of the Central Bureau of Statistics, Sweden; Mr. B. A. Liu, Chief of Statistical Division, Department of Social Sciences, UNESCO; Mr. W. F. Ogburn, Indian School of International Studies, New Delhi ; Professor E. S. Pearson, Royal Statistical Society and University College, London and Mr. G. E. Wood, Census and Statistics Department, Wellington.

The Pakistan Statistical Association sent heartiest felicitations and referred to the Indian Statistical Institute as occupying a unique position in the East and an international status in Statistics. Other congratulatory messages were received from Sri G. B. Pant, Minister of Home Affairs, Government of India; Sri J. J. Anjaria, Chief of the Resources, Economic Survey and Finance Division, Planning Commission, New Delhi; Professor Radhakamal Mukherjee, Vice-Chancellor, University of Lucknow; Dr. R. Balakrishna, Professor of Economics, University of Madras; Professor K. N. Bagchi, Principal, National Medical College, Calcutta on behalf of Indian Science Congress Association and National Medical Institute; Sri S. Subramaniam, Joint Director, Central Statistical Organization, New Delhi; Professor V. R. Pillai, University College, Trivandrum and Sri M. Verma, University Training College, Nagpur.

Appendix 3: List of Papers completed during 1956-57

A. Theoretical Statistics

A.1 SUMMARY OF RESEARCH WORK IN THEORETICAL STATISTICS

(The numbers in bracket are those of papers listed in Section A.2)

Probability Theory (7, 21, 22): It is shown that the Fourier transform of a finite probability distribution, which is an integral function, cannot be without zeros in the entire complex plane. From this, it follows as a corollary that a random variable following a proper infinitely divisible law cannot be bounded.

Given a probability space (X, S, μ) and a sequence of random vectors ξ_1, ξ_2, . . . on it, such that the induced measures $\mu\xi_1^{-1}$, $\mu\xi_2^{-1}$, . . . converge weakly to a measure P, it is proved that there is a random vector ξ, which is such that $P = \mu\xi^{-1}$.

It is proved that any discrete stochastic process can be generated from a rectangular variable. No independent family of proper random variables, however, can be obtained from the rectangular variable, if the family is uncountable.

Characterization Theorem (6): A characterization of the normal distribution is obtained as a consequence of the linearity of multiple regression and homoscedasticity of the conditional distribution, when all the variables concerned are interconnected by a linear structural relation.

Sequential Analysis (1): It has been shown that the fundamental identity of sequential analysis is a special case of a formula for the probability that the sampling terminates at some finite stage. The formula is valid for any given sampling rule and yields simple proofs of known results and also some new ones, *e.g.*, upper and lower bounds for the probability of eventual termination of a random walk in the case, when there is one fixed barrier and a draft away from the barrier.

Multivariate Analysis (14, 15): In Biometric work, functions of characters representing size and shape of an organism have been found to be extremely good discriminators between species, types and so on. New defnitions have been provided for size and shape factors, which are used in the classification of groups of persons living in different states of India. A size factor is defined as one which, when its value is increased by a unit amount, produces on the average desired amounts of increase in the individual characters.

A shape factor on the other hand increases the magnitude of some of the characters and decreases the others.

New techniques have been provided for comparing growth curves under different conditions. By a suitable transformation of the time axis, which can be computed from the *observed data themselves*, the essential comparison between groups of growth curves can be made to depend on a linear rate of growth and an exact test of significance used. The adequacy of such an analysis can also be examined by a suitable test of significance.

Design of Experiments (2, 3, 4, 5, 16, 17): An expression for the efficiency of a general two-way design is obtained in terms of the association matrix and the optimality of the balanced incomplete block designs is established.

Simplified proofs of some theorems on optimal designs are given. A set of necessary and sufficient conditions, which characterise the matrix $A'A$, where A is the design matrix, for an optimal design is derived. The criterion of optimality is based on considerations of bounds on power of testing a certain linear hypothesis or estimation of some parametric function with maximum precision or simplified analysis.

A new class of designs are introduced for fitting response surfaces. These designs have the property that the confidence interval for the optimum point obtained by fitting a lower degree polynomial remains unaffected by the fact that a certain higher degree polynomial would give an adequate fit to the response surface. Other designs are constructed to explore the response surface in the near stationary region. Some of these designs could be used in stages.

Block designs for fractional replicates for 2×3^m, 3×2^n, $3^m\times 2^n$ experiments have been constructed, using certain properties of orthogonal arrays and a theorem previously given by C. R. Rao. It has been shown that incomplete blocks may be used to have designs for the fractional replicates of factorial or partially factorial experiments.

Estimation (8, 18, 13): The investigation on the efficiency of ratio estimates and two-moments estimates, mentioned in the previous report, has been extended to fairly general class of discrete distributions. Expressions for bias in these estimates to order $1/n$ (n being the sample size) are derived. Computational methods for determining maximum likelihood estimates in these cases are developed, which require a table of values of the distribution for various values of the parameter at sufficiently close intervals. The standard error of this estimate can also be approximately evaluated by using this table.

For the same general class of discrete distribution an explicit expression for the UMVUE of the parameter has been drived which has been tabulated for small sample sizes in the particular case of Poisson distribution truncated at the origin.

Sample Surveys (9, 10, 11, 12, 20): Different types of ratio estimates, based on estimates obtained from n independent and interpenetrating sub-samples, are examined from the point of view of bias and mean square error. A correction for the bias in ratio estimates is proposed, which increases the precision of the estimates in many practical cases.

A generalised unbiased ratio estimator is considered and selection procedures providing such estimators are suggested for a variety of sampling designs.

As an extension of the *pps* sampling designs new design are suggested, which gives highly efficient estimates, even when the variate y under enquiry is approximately linear in the auxiliary variate x.

A.2. PAPERS

1. BAHADUR, RAGHU RAJ : A note on the fundamental identity of sequential analysis (submitted to *Ann. Math. Stat*).
2. CHAKRAVARTI, INDRA MOHAN : Simplified proofs of some results in the theory of optimal designs (submitted to *Sankhyā*).
3. ——— Fitting response surfaces (submitted to *Sankhyā*).
4. ——— Block designs for fractional replicates of asymmetrical factorial experiments (submitted to *Sankhyā*).
5. ——— Contributions to the design and analysis of factorial experiments (a thesis submitted to Calcutta University for D.Phil degree).
6. LAHA, RADHA GOVINDA : On some characterisation problems connected with linear structural relations (*Ann. Math. Stat.* in press).
7. ——— On a property of Fourier transforms of finite probability distributions (submitted to *Proc. Amer. Math. Soc.*).
8. MITRA, SUJIT KUMAR and J. ROY : Unbiassed minimum variance estimation in a class of discrete distributions (in press).

9. MURTHY, M. N. and N. S. NANJAMMA : Almost unbiassed ratio estimates based on interpenetrating sub-sample estimates (submitted to *Sankhyā*).

10. ——— N. S. NANJAMMA and V. K. SETHI : Some sampling systems providing unbiassed ratio estimates (submitted to *Sankhyā*).

11. NANZAMMA, N. S. and M. N. MURTHY : Almost unbiassed ratio estimate based on interpenentrating sub-sample estimates (submitted to *Sankhyā*).

12. ——— MURTHY M. N. and V. K. SETHI : Some sampling systems providing unbiassed ratio estimates (submitted to *Sankhyā*).

13. PATIL, G. P. : On problems of estimation in a class of discrete distributions (a thesis submitted for the Associateship of the Indian Statistical Institute).

14. RAO, C. RADHAKRISHNA : Bengal Anthropometric Survey, 1945 Part II : Statistical analysis (in press, *Sankhyā*).

15. ——— Statistical methods for the comparison of growth curves (submitted to *Biometrics*).

16. RAMAKRISHNA, C. S. : Some contributions to the theory of designs involving only two replications (Thesis submitted for the Associateship of the Indian Statistical Institute).

17. ROY, JOGABRATA : On the efficiency factor of incomplete blocks desgins (submitted to *Sankhyā*).

18. ——— and S. K. MITRA : Unbiassed minimum variance estimation in a class of discrete distributions (in press).

19. ROY CHOWDHURI, D. K. : On some extensions of *pps* sampling (a thesis submitted for the Associateship of the Indian Statistical Institute).

20. SETHI, V. K., N. S. NANZAMMA and M. N. MURTHY : Some sampling systems providing unbiassed ratio estimates (submitted to *Sankhyā*).

21. VARADARAJAN, V. S. : On a problem in measure-spaces (submitted to *Ann. Math. Stat.*).

22. ——— On an existence theorem in probability spaces (submitted to the *Academy of Sciences, Moscow*).

B. Applied Statistics

B.1. BIOMETRIC STUDIES

1. DAS, B. C. : Differential rate of survival of the spawn of economic fish as a function of water treatments. (Communicated to the Indian Science Congress, January 16, 1957, Psysiology Section).

2. ROY, S. K. : Studies on the activities of earthworms. (Communicated to the Indian Science Congress, January 1957).

B.2. Psychometric Studies

1. Chatterjee, S.: Standardisation of Iowa Mathematics aptitude test under Indian conditions. (Communicated to Psychology and Educational Sciences Section, Indian Science Congress Association, Jan. 1957).
2. Das, Rhea S.: Report on the IMRUP personnel selection program (rotaprinted).
3. ———: A methodological analysis of interview data. (Communicated to Psychology and Educational Sciences Section, Indian Science Congress Association, January 1957).
4. ———: The interview in personnel selection (rotaprinted).
5. ———: A syllabus for psychometrics for use in India (rotaprinted).
6. ——— (with S. K. Mitra): A report on the evaluation of the NSS (operations) training and discussion conference (rotaprinted).
7. Dasgupta, B.: A proposed weighting and scoring method in performance tests. (Communicated to Psychology and Educational Sciences Section, Indian Science Congress Association, January 1957).
8. ——— (with S. P. Sangal): Explanation of psychometric research and service chart showing the Davis difficulty and discrimination indices for item analysis and instructions for its use.
9. ———: Some aspects of scaling in the two-stage selection process. (Communicated to Statistics Section, Indian Science Congress Association, January 1957).
10. Harper, A. Edwin, Jr.: Recent advances in psychometry, Vidya Bhawan Society Silver Jubilee Vol. (in press).
11. ———: Brief summary of studies concerning the validity of the Vellore Christian Medical College selection methods. (Confidential report to the College Board).
12. ——— (with T. K. Sen): Revision of Diederich's Test of critial analysis of Reading and Writing with accompanying notes (rotaprinted).
13. ———: Booklet on methods of hand and machine scoring of tests (revised edition) (rotaprinted).
14. Mitra, Shib K. with Siske, D. W.: Intra-individual variability as related to test score and item. (Educational and Psychological Measurement, Vol. 16, No. 1. 1956, pp. 3–12).
15. ——— (with R. S. Rhea): Report on the evaluation of the NSS Operations training and discussion conference, July 1956 (rotaprinted).
16. ——— PRSU Technical Report on RTS Short-term trainees selection test, March 1956.
17. Mukherjee, S.: Empirical data on the accuracy of Jenkins' short formula for estimating sigma.
18. Sangal, S. P.: Explanation of Psychometric Research and Service Chart showing the Davis Difficulty and Discrimination Indices for item analysis, and instructions for its use.
19. Sen, Tapas Kumar: An empirical study of different methods of estimating test reliability.

Appendix 4: Scientific Inquiries (Psychometric unit)

A. Inquiries which have led to the collection and/or analysis of data according to schemes developed in consultation with staff members of PRSU.

Durganand Sinha, *Patna University :* Item analysis methods and selection of items for a non-verbal test of intelligence.

N. Sircar, *Utkal University :* The specificity of trait aggression in a questionnaire.

H. S. Gill, *East Punjab University :* Research design for the validation of some hypotheses concerning the Roschach test.

S. Pathak, *Patna University :* Analysis of data on attitudes towards marriage.

N. C. S. Rao, *Jabalpur :* Factor analysis of attitude scale items to test some of Eysenck's factors.

Mrs. Lina Ray, *Calcutta University :* Analysis of data from a test of English literary aptitude.

B. De, *L.S. College, Bihar University :* Analysis of symbol association data collected from normal, anxiety and psychotic patients.

A. N. Deb, *Ravenshaw College, Cuttack :* Beliefs in the causes and treatment of crime.

Prem Shankar, *Department of Psychology, Lucknow University :* Measuring change of attitudes in community project areas.

S. S. Mukherji, *Calcutta University :* Study of juvenile delinquency in Calcutta.

G. B. Kapat, *Department of Education, Calcutta University :* Statistical treatment of data for standardisation and validation of a scholastic aptitude test in Bengali for 5th grade school children in Bengal.

V. C. Rambo, *Mangeli and Vellore :* Statistical analysis of data on ocular convergence in India and Europe.

Anglo-Indian Survey Committee : Problems connected with a socio-economic survey of the Anglo-Indian community in Calcutta.

B. Other inquiries in which technical advice has been given by PRSU.

Durganand Sinha, *Patna University :* (i) Measurement of anxiety with questionnaires.

(ii) Reliability and validity problems in the measurement of leadership in NSS cadets with verbal group tests.

B. C. Asthana, *Aligarh University :* Scaling of differently weighted papers in an examination.

Sri Ojha, *Delhi :* Research design for the study of validity of the Children's Apperception Test.

Radhanath Rath, *Utkal University :* Attitude scale construction.

B. D. Bhatia, *College of Nursing, Delhi:* The problem of reliability and validity in a measurement of recessiveness in school children.

Miss P. Sheopuri, *Central Institute of Education, Delhi:* Effect of childrearing practices on the personality of young children as revealed through the medium of doll play.

Biswanath Mukherji, *Institute of Child Development, Ahmedabad:* Research design for the study of the stability of perceptual attitudes and its relation to certain personality factors.

A. N. Sen, *National Tobacco Co:* (i) Problems of classification and placement; (ii) Market research.

B. De, *L. S. College, Bihar University:* Development of a selection battery for police inspectors.

S. Jalota, *Banaras Hindu Univesity:* Analysis and scaling of work association data.

A. K. Pal, *Calcutta University:* Measurement of the criminal personality.

C. T. Phillip, *All India Radio:* Effect of radio on the cultural pattern of the country.

S. P. Sundar Rao, *Christian Medical College, Vellore:* Health questionnaire analysis.

Srinivas Bhattacharya, *David Hare Training College, Calcutta:* Construction and analysis of interest tests.

Appendix 5 : Workers admitted to the degree of Doctor of Philosophy

1. R. G. Laha (RTS) : Characterization of Probability distribution and statistics from suitable stochastic relations.
2. A. Matthai (RTS) : Some techniques of planning sampling investigations in Statistical Quality Control and Sample Surveys.
3. J. Roy (RTS) : On some problems in Multivariate analysis.

Workers who submitted dissertations for D. Phil degree or AISI (Associateship of the Indian Statistical Institute)

1. I. Chakravarti (RTS) for the D. Phil degree : Contributions to the Design and analysis of factorial experiments.
2. G. Hariharan (SQC) for the AISI : Report on the professional work in statistics carried out by him as a member of the SQC Unit.
3. G. P. Patil (RTS) for the AISI : On some problems of estimation in a class of discrete distributions.
4. D. K. Roy Choudhury (RTS) for the AISI : On some extensions of *pps* sampling.
5. C. S. Ramakrishnan (RTS) for the AISI : Dual of a two-associate PBIB design and contributions to designs involving only two replications, together with a list of all such designs.

Appendix 6 : International Statistical Education Centre (ISEC)

List of Trainees

Tenth Term : July 1956 to April 1957

Burma : 1. Kyaw Htay, 2. Maung Maung.

Ceylon : 3. K. N. W. Upali Godammune, 4. T. C. Perreira Gunnaratne.

India : 5. Asit Ranjan Choudhuri, 6. Prem Lal Gupta, 7. K. Krishnan Nair, 8. Aranjo Anthony, 9. M. Sankaranarayana Pillai, 10. Sudhir Kumar Banerjee, 11. Sudhir Kumar Chatterjee, 12. N. B. Bagchi, 13. Gauri Sankar Chatterjee, 14. Dibyendu Kumar Mitra.

Japan : 15. Teruo Nomura, 16. Hiroshi Sukegawa, 17. Yuichi Yoshioka, 18. Tetsuo Oguma.

Pakistan : 19. Asad Zaman, 20. Md. Auqil Seyed, 21. Md. Ataur Rahman, 22. Equb Ali Khan.

Phillippines : 23. Abadeso P. Gillera, 24. Antonio S. Alfonso.

Singapore : 25. Chua Yew Keng, 26. Miss Sally Yeo.

Thailand : Virool Boonyasombut, 28. Miss Suwatana Suksmarn.

Appendix 7 : List of successful candidates in professional examinations

Statisticians' Diploma Examination : September, 1956

A. General Papers

Paper I (Theoretical) : Avelino Jose Romandus Gonsalves (B2), Narayan Mahabaleswar Palekar (B6), Ramesh Laksiman Kamat (B9), Sashikant Kashinath Gadgil (B10), Meghasham Bhaskar Usagaunkar (B11), Vasudeo V. Ghalsasi (B17), Jagdish Jha (C1), Sukumar Chakravarti (C10), Nirmal Kanti Das Gupta (C11), Nirmalendu Bhowmick (C23), Sukhendu Nath (C24), Tapeswari Kumar Trivedi (D1), Avtar Singh Chawla (D2), Y. Shiva Rao (D12), Krishan Gopal (D14), Gopal Dev Ahluwalia (D16), Sudhindra Nath Ganguly (D22), Surendra Kumar (D26), S. Harihara Iyer (D34), Brij Mohan Mahajan (D47), Venkata Subramanian Radhakrishnan (M2), Arun Prabhakar Kulkarni (P6).

Paper II (Theoretical) : Shankar Shripad Telang (B5), Jagdish Jha (C1), Ela Romola Mukherjee (C3), Kalyan Das Gupta (C8), Prakash Chandra Kundu (C22), Nirmalendu Bhowmik (C23), Sukhendu Nath De (C24), Shankar Bhaduri (C25), Manoj Chatterjee (C27), Ved Prakash Agarwal (D3), Dhara Puram Sundaresaiyar Ramaratnam (D27), Rajeswar Dayal Saxena (D28), Balaram Kohli (D48), Jugal Kishore Sharma (L1).

Paper VI (Practical) : Dipak Chakravarti (C9), Gour Gopal Rudra (C15), Gopal Dev Ahluwalia (D16), Shriram Krishna Karandikar (P5), Arun Prabhakar Kulkarni (P6).

Paper VII (Practical) : Shankar Bhaduri (C25).

B. Special Papers

Papers IV and V (Theoretical)

(1) *Sample Survey (Applied) :* Srish Chandra Baswria (C7), Amalendu Sengupta (C13), Nirmalendu Bhowmik (C23), Tilak Raj Talwar (D25), Dev Raj Chawla (D32), Bharat Bhusan Behl (D42).

(2) *Sample Survey (Theoretical) :* Ved Prakash Agarwal (D3), S. Harihara Iyer (D34), Jugal Kishore Sharma (L1), Jagdish Narayan Srivastava (L4).

(3) *Economic Statistics :* Tilak Raj Talwar (D25).

(4) *Statistical Quality Control :* Sukumar Chakravarti (C10), Nirmal Kanti Dasgupta (C11), Amalendu Sengupta (C13), Samir Ranjan Guha Roy (C18), Nirmalendu Bhowmik (C23), Sashi Kant Mathure, Gopal Dev Ahluwalia (D16).

(5) *Design of Experiment* (*Applied*) *:* Narindar Kumar Sahni (D18), Jagdish Kumar (D36), Brij Mohan Mahajan (D47), Jugal Kishore Sharma (L1).

(6) *Mathematical Theory of Sampling Distribution :* Venkata Subramanian Radhakrishnan (M2), Arwind Parasuram Joag (P14).

Papers VIII and IX (Practical)

(1) *Sample Survey* (*Applied*) *:* Kalyan Dasgupta (C8).

(2) *Sample Survey* (*Theoretical*) *:* Ved Prakash Agarwal (D3), Niranjan Singh (D28), S. Harihara Iyer (D34), Nurani Parameswara Ayyar Mahadevan (M6).

(3) *Statistical Quality Control :* Sukumar Chakravarti (C10), Sashi Kant Mathure (B7), Vasudeo V. Ghalsasi (B17), Gopal Dev Ahluwalia (D16).

(4) *Design of Experiment* (*Applied*) *:* Avtar Singh Chawla (D2), Niranjan Singh (D28), Tarun Kumar Gupta (D46).

(5) *Mathematical Theory of Sampling Distribution :* Venkata Subramanian Radhakrishnan (M2).

Computer's Certificate Examination : September, 1956

Part IA, Sec. I : Kunja Behari Ghosh (C1), Nirmalendu Basu Chaudhury (C4), Amiya Kishore Dasupta (C5), Akhilendu Dutta (C7), Nirmal Krishna Chaudhury (C10), Madhab Munshi (C11), Sudhangsu Bhusan Das (C12), Sanat Kumar Banerjee (C14), Manindra Chandra Basu (C15), Chittaranjan Chatterjee (C16), Amal Kanti Bhattacharyya (C20), Tushar Kanti Banerjee (C25), Ranjan Kumar Bhattacharyya (C30), Kamalendra Nath Duttarai (C31), Chunilal Basu Chowdhury (C32), Raj Kumar Nandy (C33), Amarendra Nath Ghosh (C38), Samar Krishna Basu (C46), Amiya Bhusan Chakravarti (C49), Animesh Chakravarti (C64), Ramchandra Mitra (C67), Amal Kumar Banerjee (C71), Biman Chandra Mitra (C73), Amiya Kumar Sinha (C79), Dulal Kanti Choudhury (C80), Sambhu Nath Bhattacharjee (C83), Syamal Kumar Sen (C86), Amal Roy Choudhury (C87), Patit Paban Banerjee (C91), Paresh Nath Bhattacharjee (C97), Nirmal Kumar Choudhury (C99), Nagendra Chakravarti (C114), Nirmal Chakravarti (C117), Nirendra Nath Sarkar (C119), Asoke Kumar Choudhury (C121), Devendra Kumar Debnath (C122), Dilip Kumar Ghosh (C130), Sachi Ranjan Ganguly (C133), Subhrangsu Bhattacharyya (C134), Tapan Dasgupta (C135), Anil Kumar Singha (C138), Ananta Mohan Dey (C154), Sudhindra Sarkar (C155), Ranjit Chakravarti (C161), Parimal Mukherjee (C163), Keshab Chandra Das (C166), Balai Chandra Das (C167), Niraj Kanti Das (C169), Narayan Chandra Sen (C173), Sisir Kumar Ghosh (C176), Arun Kumar Das (C180), Marud Baran Chakravarti (C184), Mrinal Kanti Chatterjee (C187), Madhusudan Dutta (C192), Santosh Chandra Kundu (C198), Shanti Priya Bhowmik (C200), Anil Kumar Sengupta (C203), Haridas Roy (C204), Bharati Chakravarti (C207), Shib Ranjan Banerjee (C211), Arun Sanyal (C214), Jyotirmoy Basak (C215), Ramesh Chandra Jeirath (D3), Dinabandhu Das (G8), Md. Muslim (G13), Giridhar Chatterjee (G14), Khanweldar Madhusudan Ramachandra (P4).

Part IA, Sec. II : Akhilendu Dutta (C7), Mrinal Kanti Roy (C9), Nirmal Krishna Choudhury (C10), Madhav Munshi (C11), Manindra Chandra Basu (C15), Amal Kanti Bhattacharyya (C20), Tushar Kanti Banerjee (C25), Anil Kumar Bhattacharyya (C27), Hari Bhusan Dutta (C28), Chunilal Basu Chowdhury (C32), Samarendra Barua (C47), Pranabananda Bhaduri (C50), Ramchandra Mitra (C67), Syamal Kumar Sen (C86), Paresh Nath Bhattacharjee (C97), Nirmal Kumar Choudhury (C99), Mrinal Kanti Sarkar (C107), Sourindra Nath Pal (C108), Sushil Kumar Tapadar (C116), Nirendra Nath Sarkar (C119), Keshab Chandra Banerjee (C120), Asoke Kumar Choudhury (C121), Devendra Kumar Debnath (C122), Nagendra Nath Das (C129), Ajit Kumar Sengupta (C137), Arun Prosad Sinha (C146), Tarani Kanti Palrai (C147), Monoranjan Das (C150), Narayan Chandra Sen (C173), Ajoy Kumar Sarkar (C186), Animesh Singhaw (C201), Kalidas Neogy (C205), Sanat Kumar Das (C209), Debabrata Bardhan (G1), Syed Resalat Kussain (G2), Sakti Kumar Chowdhury (G6), Md. Muslim (G13), Hari Narayan Joshi (P2), Khawelkar Madhusudan Ramachandra (P4).

Part IB, Sec. I : Madhab Munshi (C11), Ranajit Kumar Naha (C52), Pranab Kumar Mitra (C53), Sankar Bhaduri (C85), Nirendra Nath Sankar (C119), Nirmalananda Banerjee (C126), Sanat Kumar Das (C209).

Part IB, Sec. II : Amiya Kishore Dasgupta (C5), Madhab Munshi (C11), Tushar Kanti Banerjee (C25), Radhashyam Nath (C36), Subir Kumar Roy (C37), Amarendra Nath Ghosh (C38), Subodh Kumar Pal (C39), Lakshmi Narayan Das (C42), Nihar Ranjan Saha (C45), Sanat Kumar Bhose (C48), Pranabananda Bhaduri (C50), Loknath Mukherjee (C51), Ranajit Kumar Naha (C52), Asoke Dasgupta (C58), Monoj Chatterjee (C59), Kartik Chandra Bhattacharjee (C60), Biswapati Mukherjee (C72), Dulal Kanti Choudhury (C80), Sambhu Nath Bhattacharjee (C83), Prafulla Kumar Basak (C84), Tinkari Pal (C100), Arun Kumar Mitra (C101), Manik Ratan Acharyya (C106), Sibakar Ghosh (C115), Sushil Kumar Tapadar (C116), Amalendu Sengupta (C125), Rabindra Kumar Chakravarti (C127), Prasanta Kumar Sinha (C128), Nagendra Nath Das (C129), Arun Chandra Chakravarti (C131), Probhas Mukherjee (C132), Ajit Kumar Sengupta (C137), Pranesh Chandra Kar (C140), Biswadhan Chattopadhayaya (C151), Parimal Mukherjee (C163), Keshab Chandra Das (C166), Bimalendu Mahalanobis (C175), Mihir Kumar Rakshit (C178), Rabindra Nath Mukherjee (C183), Ajoy Kumar Sarkar (C186), Srish Chandra Basurai (C195), Sukha Sinchan Roy (C196), Kalidas Neogy (C205), Sanat Kumar Das (C209), Bimal Jyoti Sanyal (C212), Sharat Kumar Mitra (G4), Subodh Kumar Palit (G17), Md. Yusuf Ansari (G23), Hari Narayan Joshi (P2).

Part IC, Sec. I : Benoyendra Goswami (C35), Ranajit Kumar Naha (C52), Sukhendu Maitra (C56), Anu Ranjan Mukherjee (C61), Subhendara Sekhar Das (C66), Nirmalananda Banerjee (C126), Rabindra Kumar Chakravarti (C127), Ajit Kumar Sen Gupta (C137), Arun Kanti Ghosal (C210).

Part IC, Sec. II : Benoyendra Goswami (C35), Nihar Ranjan Saha (C45), Sanat Kumar Bhose (C48), Pranab Kumar Mitra (C53), Asoke Dasgupta (C58), Monoj Chatterjee (C59), Dhruba Ranjan Chakravarti (C102), Manik Ratan Acharyya (C106), Nirmal Chakravarti (C117), Nirmalananda Banerjee (C126), Rabindra Kumar Chakravarti (C127), Prasanta Kumar Sinha (C128), Tapan Dasgupta (C135), Ajit Kumar Sengupta (C137), Biswadhan Chattopadhayaya (C151), Gaur Chandra Mukherjee (2152), Swaraj Kanta Pal (C159), Kulada Ranjan Biswas (C170), Badal Basu Mullick (C194), Srish Chandra Basurai (C195), Arun Kanti Ghosal (C210), Kalidas Chatterjee (C216).

Statistical Field Survey Certificate Examination : September, 1956

Part I, Sec. A : Venkataraman Srinivasarao (B8), Chitta Ranjan Saha Roy (C13), Bharat Bhusan Behl (D4), Shyamlal Aggarwal (D5), Pyare Lal Sharma (D6), Surinder Singh Kapoor (D12), Basudeb Prosad (G2), Ramachandra Prosad (G7), Thakur Indradeo Sharma (G9), Narinder Singh Sachar (L3), Sidhanand Nathani (L5).

Part I, Sec. B : Sudhi Ranjan Dasgupta (C5), Bharat Bhusan Behl (D4), Shyamlal Aggarwal (D5), Pyare Lal Sharma (D6), Balbir Singh Bakshi (D11), Md. Muslim (G4), Narinder Singh Sachar (L3), Sidhanand Nathani (L5).

Part I, Sec. C : Lalit Mohan Chatterjee (C3), Bharat Bhusan Behl (D4), Shyamlal Aggarwal (D5), Pyare Lal Sharma (D6), Ranjit Mukherjee (G5), Thakur Indradeo Sharma (G9), Narinder Singh Sachar (L3), Sidhanand Nathani (L5), Dori Lal Pachauri (L14).

Part II, Sec. A : Kamalendra Nath Duttarai (C6), Jahar Lal Ghosh (C14), K. Venugopal (D3), Adisubramanya (B4), J. R. Doraiswamy Iyer (B5), Pyare Lal Sharma (D6), Sidhanand Nathani (L5).

Part II, Sec. B : K. Venugopal (D3), Sidhanand Nathani (L5).

Appendix 8 : NSS Draft Reports and Working Papers

DRAFT REPORTS

1. NSS No. 13 : Tables with Notes on the fourth round (April-September 1952) Vols. 1 and 2.
2. ,, 14 : Some characteristics of the economically active population.
3. ,, 15 : Report on sample-survey of Employment and Unemployment (May-November 1955).
4. ,, 16 : A Technical note on Age Grouping (4/R, West Bengal Sp. Study, 1954 E and ESP Study, h-h. comparative, 1955).
5. ,, 17 : A Report on Small Scale Transport Operation 7th round (October 1953-March 1954).
6. ,, 18 : Sample Verification of Livestock Census, 1956.
7. ,, 19 : Pattern of Consumer Expenditure, 2nd to 7th rounds, (April 1951-March 1954)
8. ,, 20 : Small Scale Manufacture (October 1953-March 1954, Schedule 2.2, 7/R).
9. ,, 21 : Small Scale Manufacturing Establishments (9/R), May-November 1955.
10. ,, 22 : Notes on Investigators' Time Records (8/R), July 1954-April 1955.

WORKING PAPERS

1. A note on the nature of the basic data on large industrial establishments in India.
2. Evaluation of the N.S.S. 11th round Training and discussion conference of July 1956.
3. Growth of urban Delhi.

Appendix 9 : Studies under Research Unit, NSS

1. DESIGN UNIT : *Sampling Design :* (i) Almost unbiassed ratio estimates based on inter-penetrating subsample estimates. (ii) Some sampling systems providing unbiassed ratio estimates. (iii) Ordered and unordered estimators in sampling without replacement. (iv) Some experiments in the designing of crop acreage surveys. (v) Balanced systematic sampling. (vi) Some aspects of sampling theory and practice.

2. ROY, J. : *Review of old data :* Note on inconsistent entries, non-entries etc. relating to employment and unemployment enquiry (sch. 10, 9th round, bls. 9 and 10).

2.1. HALDER, S. : Note on occupational distribution of W. Bengal urban population (sch. 10, 9th round, W. Bengal urban, subsamples 1 and 2).

2.2. BANERJEE, S. : (i) Quantitative estimates of consumption of cereals (7th round, rural, SS–1 and 2). (ii) Some estimates of sampling error in consumer expenditure estimates. (iii) Note on data on consumer expenditure and reference period.

2.3. DEY, B. : (i) Quantitative estimates of consumption of cereals (5-6th rounds, rural SS–1 and 2). (ii) Mistakes in the nature of omission and commission in Employment and Unemployment Enquiry (sch. 10, 9th round, bl. 13).

2.4. BANERJEE, U. N. : A note on reliability of gross profit estimate from household trade schedule. (sch. 2.4, 7th round, rural). (ii) Notes relating to (a) household trade concept and definition, (b) blank schedules and blank entries—extent of non-response, (c) wholesale trade, (d) retail trade, (e) gross value of purchase in rural retail trade (sch. 2.4, 7th round, rural, pilot SS–1).

2.5. BHATTACHARYA, S. : (i) A note on handloom weaving industry in W. Bengal and Madras with special reference to cost in relation to receipts (sch. 2.2, 7th round, rural, SS–1 and 2). (ii) A note on comparison of certain estimates as compiled from NSS household manufacture enquiry (sch. 2.2, 7th round, SS–1 and 2) and as given in the Economic Survey of Small Industries of Calcutta for 1952–53 undertaken by West Bengal State Statistical Bureau.

2.6. BHATTACHARYA, S. : A note on some primary aspects of Sample Survey of Manufacturing Industries (Source material : Survey of Company Finance published in the Reserve Bank of India Bulletin in July, 1955).

3. BANERJEE, U. N. : *Review of current data :* Notes in the nature of queries, clarifications and suggestions for drawing the attention of the Field, based on scrutiny of data in agr. labour enquiry (sch. 10.1, 11th round, rural SS–1 and 6).

3.1. BANERJEE, S. : Scrutiny notes in connection with consumer expenditure enquiry (sch. 1.1, 11th round, rural).

4. MITRA, C. : *Field Studies :* (i) Analysis and Report on Pilot Survey No. 1 (A study designed to have an idea of response error and means to control it. (ii) Analysis and Report on Pilot Survey No. 3 (A study on nature and extent of response discrepancies to the same set of questions put by different interviewers within a short time interval through detailed and condensed schedules).

4.1. Banerjee, U. N. : A study of investigator's time-record (on the basis of data provided in pilot survey No. 2 conducted in 30 villages of Hooghly District in West Bengal).

5. Ghosh, N. C. : *Special Studies :* A note on per capita cloth consumption in India in 1953–54.

5.1. Banerjee, U. N. : A note on validity of age data (on the basis of data collected in pilot survey No. 2).

5.2. Halder, S. : A note on comparison of basic materials for NSS sampling frame (tehsils) with published census data— for Assam.

5.3. Dey, B. : A note on comparison of basic materials for NSS sampling frame (tehsils) with published census data— for Orissa.

5.4. Banerjee, U. N. : A note on comparison of basic materials for NSS sampling frame (tehsils) with published census data— for West Bengal.

Appendix 10 : Studies under Operational Unit

1. *Experiments on land utilization surveys in cadastrally unsurveyed areas, Giridih (August–October 1956) :* Experiments on 'line sampling' in land utilization surveys without making use of the village cadastral survey maps were carried out at Giridih. The investigator was required to locate sample points at random over an area of 225 square miles guided duly by a magnetic compass and a Survey of India map drawn to a small scale. At each point he was required to enumerate line units of varying length. Similar experiments on a larger scale covering 4000 square miles were subsequently completed in the district of Hazaribagh. A number of studies was also undertaken to locate random points and enumerate line units within villages, for which no maps are available, either showing the plot boundary or showing the village boundary. On the whole, the result of their studies were encouraging.

2. A study of the distribution of harvestable fields on different days of the crop season was undertaken during August–October 1956 in three centres of West Bengal. Unusual rains, however, partly affected the quality of data collected.

3. *Special crop-cutting experiments :* Sample cuts of various shapes and sizes were tried out in Giridih on *Aman* paddy. The experiment on the estimation of paddy yield by progressive harvesting was continued from the previous year.

4. A small experiment was carried out in two centres to study the weight of field enumeration and variability for certain selected items of household consumption using varying lengths of reference periods.

5. *Experiments with 'motorized units' for conducting crop-cutting experiments :* The entire district of Birbhum in West Bengal was covered by the survey, in which the investigation was conducted by three teams, the first working on foot and covering 12 villages in a month and the third working with a motor car and covering 24 villages in a fortnight. The results of the survey are likely to be of interest for planning future programmes.

6. *Miscellaneous surveys :* Various other surveys were conducted by the special unit and mention may be made of the following : A revisit of 410 households in Calcutta

covered in an earlier survey (April–September 1952) was organized with an aim to study the change in consumption pattern. Only 260 of the households could be traced in the revisit. Another study was made again in Calcutta to find out how the information (on consumer expenditure) collected varies, when different persons in the household happen to be the respondents. A similar study was made in respect of opinion on certain items of common interest to find out how opinions vary between informants in the same household. Finally, a study was made on the basis of past data to find out whether the crop yield rate had any relation with the distance of the plot from the nearest road. In case there was no association, a cheaper crop-cutting survey would be possible by using plots adjoining roads.

Appendix 11 : Other projects and Studies

1. *Experiments on different methods of paddy cultivation :* Ten experiments were conducted in Giridih on *Aus* and *Aman* crops to study the effects of different methods of paddy cultivation. In particular, interactions between spacing and number of seedlings per hold and the effects of different kinds of fertilizers were observed in these experiments.

2. *Progressive Maturity of Winter Paddy :* Another experiment in Giridih related to a study of progressive maturity of *Aman* paddy transplanted according to the Japanese method. Altogether 3000 samples were collected and the analysis is in progress.

3. *Effect of manure on sugarcane :* A third experiment in Giridih related to a study of the effect of manure on different methods of sugarcane cultivation. The experiment incorporated observations on the ratios between weights of sugarcane and juice, bagasse and *gur* for crops grown by different methods of cultivation.

4. *Family Planning survey :* A preliminary survey was undertaken in 49 census blocks of Calcutta to study family planning. Married persons in 8000 households were interrogated.

5. *Special health survey :* A somewhat similar survey related to the study of benefits of hospital services in West Bengal. The survey involved contacting a sample of ex-patients of hospitals.

6. *Experimental study of trading in West Bengal :* This survey was started in January 1957 with the object of studying the level and character of trading activity in urban West Bengal and incidentally improving the trade schedule in NSS general enquiry. This was a fairly large survey involving 10 investigators under two inspectors.

7. *Survey with pre-coded schedule and mark-sensing cards :* A pilot survey was carried out during winter 1957 with pre-coded schedules. The figures, which related to four villages, were then transferred to mark-sensing cards.

8. *Verification of census enumeration work :* A complete enumeration of population in census slips was taken up in the middle of November 1956. The same operation was repeated in another winter in the same area, as soon as the work of the first worker was completed. In all three census blocks were covered in the study and the analysis of data is nearly over.

9. *Special enquiry on consumption :* A special study on consumption in Calcutta was initiated from January 1957. The work, which concerns study of different biases, is continuing.

Appendix 12 : Working Papers on Planning, 1956-57

57. Roy, Ajit and Banerjee, Nripen : Monopolies and concentration of economic power—Part I. (24 April 1956).
58. Upadhyaya, P. K. : Measurement of the density of Railway Passenger Traffic. (28 May 1956).
59. Sengupta, S. S. : Rates of utilization in selected manufacturing industries. (18 June 1956).
60. Narasinham, G. V. L. : A note on the utility of Cobb-Douglas type of production function for Indian Manufacturing Industries. (28 June 1956).
61. Roy, Ajit : Indian Labour during First Five Year Plan. (2 July 1956).
62. Bettelheim, Ch. : The technological choices and the aim of economic planning long-term planning : Paper No. 1. (9 July 1956).
63. Chatterjee, N. K. : Shipping foreign and exchange. (12 July 1956).
64. Sharma, I. R. K. : Relation between investment and production of cement and steel in some selected countries. (14 July 1956).
65. Chatterjee, N. K. : Non-essential imports and foreign exchange. (15 July 1956).
66. Sengupta, S. S., Chakravarty, A. and Agrawal, V. K. : A note on forecasting net product of 1953–54. (17 July 1956).
67. Sengupta, S. S. : A model for periodical planning. (20 July 1956).
68. Bettelheim, Ch. : Basic data needed for solving the problems of investment allocation and of technological choices : long-term planning : Paper No. 2. (21 July 1956).
69. ——— : The allocation of investments : long-term planning ; Paper No. 3. (23 July 1956)
70. ——— : The problem of technological choices: long-term planning : Paper No. 4. (26 July 1956).
71. ——— : Employment problems and economic growth long-term planning : Paper No. 5. (27 July 1956).
72. ——— : Modernization and Rationalization investments : long-term planning : Paper No. 6. (28 July 1956).
73. ——— : Some basic long-term planning problems—general conclusions : long-term planning : Paper No. 7. (30 July 1956).
74. Naqvi, K. A. : The problems of the elaboration of the provisional project of a long-term plan of India: Paper No. 8. (31 July 1956).
75. ——— : The methodology of the elaboration of the provisional project of a long-term plan of India: Paper No. 9. (1 August 1956).
76. ——— : The problems in the choice of technique during the Fifteen Year plan (1956–57—1970–71) of India: Paper No. 10. (2 August 1956).
77. Chakravarty, A. K. : On factors of production of coal. (4 August 1956).
78. Narasinham, G. V. L. : Estimates of material production in India from 1949–50 to 1952–53. (4 August 1956).
79. Sharma, I. R. K. : Relation between national income and production of some selected industries. (6 August 1956).

80. Joint Paper of Section III : Annual Plan for 1957–58: A Draft Outline. (9 August 1956).

81. Talwar, T. R. : Commodity balances for the basic raw material (Indian industries 1951–55). (10 August 1956).

82. Lubell, H. and Kukreja, O. P.: A graphical increase in the use of consumption patterns. (17 August 1956).

83. Rao, P. R. : Agricultural forecasts for 1955–56 to 1960–61. (24 August 1955).

84. Chatterjee, N. K. : Un-utilised capacity, imports and foreign exchange. (27 August 1956).

85. Rudra, A. and Roy, Bina : Some results in consumption studies. (7 September 1956).

86. Narsu, A. V. : Size efficiency in Manufacturing industry. (19 September 1956).

87. Ghose, M. N. : An approach to the problem of investment in Railways and road transport during the Second Five Year Plan. (19 November 1956).

88. Chatterjee, A. K. : Trends in Labour Productivity and Wages in Twelve selected industries (1946–1952). (4 December 1956).

89. Das, P. : Progress of the industrial sector (organized) during the First Five Year Plan : an assessment. (7 January 1957).

Appendix 13 : Machine Tabulation Unit

1. Strength of Tabulation Equipments at different centres

machine	Baranagar				Giridih	Delh		total
	Hollerith	IBM	Russian	Powers-Samas	Hollerith	Hollerith	IBM	
			April to September 1956					
accounting machine	5	5	–	2	1	1	1	15
statistical machine	–	2	–	–	–	–	–	2
sorter	5	8	–	2	2	2	1	20
calculating punch	–	2	–	1	1	–	1	5
collator	–	2	–	1	1	–	–	5
reproducer, gang punch & summary punch	6	6	–	3	2	2	1	20
			October 1956 to March 1957					
accounting machine	6	5	2	2	2	1	2	20
statistical machine	–	2	–	–	–	–	–	2
sorter	6	8	4	2	4	3	2	29
calculating punch	–	3	–	1	1	–	1	6
collator	–	2	–	1	1	–	–	4
reproducer, gang punch & summary punch	6	6	2	2	4	2	2	24

2. *The Output :* The per hour output from tabulating units recorded an increase from 1.39 to 2.11 thousands. The total card passage through all the tabulating units and ESM's during the year was 42660 thousands and 5623 thousands respectively. The tabulating output as recorded is the highest since the inception of machine tabulation unit.

3. Cardstores were located at Baranagar, Giridih and Delhi with 400 lakhs of cards stored in cabinets with 50 and 25 drawers each. This constitutes the punched cards of 2nd to 11th rounds of NSS and related surveys.

Appendix 14 : Library

1. The Central Library was concentrated in the main premises at 203 Barrackpore Trunk Road after almost complete transfer of material and personnel from 204 Barrackpore Trunk Road, where it had been partially accommodated. The Library now contains about 60,000 books and monographs besides a large collection of special material. Service centres were maintained as before at the City Office at 9B Esplanade East, Calcutta, and at Giridih.

2. *Books :* The Library acquired 2873 volumes of books against 2717 last year. Of these 624 were received as gifts and 97 in exchange. In addition to these 683 volumes from the late J. C. Sinha's collection, which was purchased year before last, were integrated into the Library stock. The total addition to the book stock was thus 3556 volumes. The arrearage of volumes awaiting processing at the end of the year was over 1700. This was mainly due to the loss of three trained and experienced workers through resignation and transfer from the processing unit, which has been lying practically in deadlock since the month of December 1956.

3. *Periodicals :* The Library received 1297 periodicals and annuals against 1190 last year. Of these 296 periodicals were subscribed to, 402 were received as gifts and 599 were received on exchange basis. This necessitated the registration and processing of 9750 serial pieces during the year, an increase of 10 percent over the previous year. The Library subscribed to 39 new journals and entered into exchange arrangements with 7 new Indian and 19 foreign agencies, among which were 3 each in Japan and USA, 2 each in China, Netherlands and Pakistan, and 1 each in Bulgaria, Hungary, Indonesia, Italy, Malaya, Poland and USSR.

4. *Special Gifts :* The most notable one was a generous gift of books and periodicals on Industrial Management valued at $ 2750 from the United States Technical Cooperation Mission in India, which was thankfully received by the Library.

For the third year in succession the Library thankfully accepted a concession of $ 125 offered by the Joint Committee on Slavic Studies (American Council of Learned Societies and the Social Science Research Council) towards the annual subscription of $ 150 for the Current Digest of the Soviet Press.

A large number of available volumes and issues of 'Wheat Studies' and several monographs of War and Peace Pamphlets, Fats and Oil studies and Grain economics were thankfully received from the Food Research Institute of the Stanford University, California.

The Library also thankfully accepted a generous gift from the Bengal Chamber of Commerce, Calcutta, comprising back volumes of several valuable periodicals, official publications, newspapers, directories, etc.

5. *Bibliographical Services :* The Library compiled three reading lists from materials available in the Library on Linear Programming, Automation and Operational Research in railway transportation.

The Library continued to issue the Weekly List of Selected Periodicals and the issue of the Index to Current Periodicals has been received since the month of August 1956. The issue of the Monthly Bulletin of New Acquisitions had to be kept suspended since the month of December 1956 in the absence of normal functioning of the processing unit for want of technical staff.

6. *Service and Circulation :* The number of persons using the Library at the end of the year was 995 against 758 last year. The total number of books, journals and other materials issued was 46271 against 38055 last year, of which 10214 were issued from the Lending Section and 36057 from the Reference Section, the number of bound volumes of periodicals, which are issued for reference only, being 14360. The total number of requests received was 54160, so that nearly 14.5 percent of the requests could not be fulfilled, against 11.5 percent last year. The rise in the number of unfulfilled requests was mainly due to the increased demand for text books from the increased number of students.

7. *Circulating Library :* The number of workers availing themselves of this Library was 764. There was new acquisition of 704 Bengali, 110 English, 72 Hindi and 43 Oriya books, bringing the total to 7510 volumes. Stocks were rotated amongst the branches of the Library. The frequency of rotation was less than in previous years owing to the duplication of the more popular books, for which there was great demand at different branches. The number of books issued from Baranagar and Giridih were 16989 and 4204 respectively.

8. *News Clippings :* The systematic processing and indexing of relevant topics relating to statistics and planning in India, which appeared in important dailies and periodicals, was continued. Clippings were gathered from 5 dailies of Calcutta and the cuttings received through the All-India Press Cutting Service, which covered more than 100 dailies and periodicals all over India. The number of clippings processed and properly indexed during the year was 8971, bringing the total number of clippings indexed so far to 20570. Articles indexed from newspapers including supplement copies numbered 1768. The indexing of articles from periodicals had to be discontinued since September 1956 for shortage of staff and the number of such articles indexed was only 377.

9. *Translation :* Several requests were received and complied with for English translation of various matters in other foreign languages, specially Russian, German and French, covering more than 300,000 words, which mostly related to economics of planning.

A Russian language class was conducted under the auspices of the Library for training up personnel to assist in translation work. Five students from different sections completed the primary course.

10. *Special Collections :* Systematic processing of the Russian books received in the Library was taken up and a classified inventory of the same containing about 500 items has been prepared.

Organization of the photo-duplicated material in the Library which required special cataloguing procedure was also taken up. The present stock comprises 226 reels of microfilm running over 2000 feet containing 21,065 exposures and 5840 pages of paper prints.

11. *Records Unit :* The Project Unit has mostly attached miscellaneous files and papers to the existing classified ones and has also arranged, classified, indexed on the 'Kardex' and made easy for reference 251 files, bringing the total number of files thus arranged to 12690 (against 12439). The Unit has received Duplicate Schedules of NSS 4th to 8th Rounds, RTS Examination papers and Crop Survey Records from Giridih for classification and processing.

In the Map Unit, 2863 West Bengal Cadastral Survey (Coloured), 1,58,800 East Pakistan Cadastral Survey, 78 Districts of East Pakistan and West Bengal and 2023 East Pakistan P.S. maps, totalling 1,63,744 (against 90,668), maps have been finally processed and shelved during the year through serial sorting and numbering, thus bringing the total number of all maps to 2,54,432 sheets. The number of maps issued to Field Branch and other unit was 29,137.

12. *Photographic Unit :* Major work on documentary reproduction consisted of 4501 frames of microfilm, 4533 paper prints from microfilm and 3491 photostats. The Photographic Unit also took 1890 still photographs of individuals, groups and important functions at the Institute as well as 3100 feet of motion picture, more than half of which was in colour. In addition to these 3004 bromide enlargements were made. Other types of work included the preparation of acre-plates, slides, photo-reproductions, hand-finished reproductions, etc. Requests from other Institutions for documentary reproduction were also complied with by supplying microfilms, photoprints and photostats. The Photographic Unit also gave training in photography to some of the Field Branch workers.

Appendix 15 : Brief Report on Visit to Japan and China : 1956

In June 1956, a team comprised of Pitambar Pant, Private Secretary to the Chairman, Planning Commission and Joint Secretary, Indian Statistical Institute, R. Natarajan, Director, IMRUP, Indian Statistical Institute and M. Mukherjee, Deputy Director, Central Statistical Organization was deputed by the Government of India to accompany Professor P. C. Mahalanobis to study planning and economic development in China and Japan.

Professor Mahalanobis, who was to have led the team and who was expected to join it in Japan on his way back from the USA, unfortunately fell ill in the USA and was advised to return to India, cancelling his projected visit to Japan and China. In his absence, Pitambar Pant acted as the leader of the team.

The team left Calcutta on 10 June 1956 and returned to Calcutta on 19 August 1956. It spent about a month each in Japan and China.

Visit to Japan

The chief interest of the team in Japan was the study of economic development in Japan, covering aspects such as :—

(*a*) organization of small scale production and its relationship to large scale factories;

(*b*) the development of machine-building and basic industries and expansion of capital goods;

(*c*) industrial management with special reference to productivity;

(*d*) salary and wage structure in industries and other occupations;

(*e*) arrangements for scientific and technical training ;

(*f*) growth and structure of capital formation in relation to the expansion of national economy and increase of national income.

The Indian Ambassador in Tokyo had intimated the Ministry of Foreign Affairs, Japan of the purpose of the visit of the team and had requested that appropriate advice and facilities be given to the team to enable it to make the best use of its time in Japan. The Ministry of Foreign Affairs entrusted Asia Kyokai (Society for Economic Cooperation in Asia) with the task of rendering advice and assistance to the team in drawing up its programme of discussions and visit. The services of Mr. Teru Sasaki, a member of the Technical Cooperation Division of the Society, were made available to the team. The members of the team were grateful to the authorities in Japan for the facilities provided to the team to have discussions with officials and non-officials, visit institutions and places of interest and collect material and establish valuable contacts. The team spent most of the time in Tokyo but were able to find a week or so for visiting industrial establishments and other institutions of interest in Osaka and Kyoto.

Within a short time of their visit in Japan, the team realized that there was a great deal of valuable material bearing on the subjects under study, but most of this material was in the Japanese language and had not been translated into English. It was obvious that in order to be able to make full use of the knowledge and experience available in Japan, it would be necessary to organize systematically the translation into English of material selected after suitable screening in consultation with Japanese experts familiar with any particular field. The team felt convinced that brief, sporadic visits of teams of experts from India to Japan to study developments in particular fields, for example small-scale industries, would in the long run be more costly and less productive of results than the organization on a continuing basis of a Unit for the translation and interpretation of selected Japanese material in different fields of interest, if such a Unit could be established in Japan under the auspices of some good institution, which should be in a position to mobilize and enlist the full cooperation of government officials, industrialists and businessmen and individual scholars.

As a result of several discussions with officials in government departments, visits to universities, research organizations, industrial enterprizes and associations and with the help of an *ad hoc* unit set up with Mr. Teru Sasaki as its leader, the team was able to collect about five hundred valuable publications, notes, memoranda, books etc., in both printed and cyclostyled forms. A list of these documents was prepared and on the advice of experts, a tentative selection was made of material, that would be worthwhile to translate from Japanese to English language for the use of planning organizations in India.

The question of establishing a Japan Reference and Translation Unit was discussed informally with several officials, individual scholars, businessmen and industrialists in Japan and the suggestion seemed to find favour with them. The matter was discussed with the Indian Ambassador and he gave enthusiastic support to the idea. Subsequently, the Asia Kyokai was discussed with the Board of Directors of that Society. The Board expressed its appreciation of the proposal, found that the function proposed was in accordance with the Society's own responsibilities and agreed to establish a Reference and Translation Unit, if Government of India so desired.

The Cabinet Secretariat was approached with a request to give approval to the project and to sanction an amount of Rupees five thousand per month for the expenses of the Unit. Necessary sanction was received, even while the team was in Japan.

The Translation and Reference Unit was established in Asia Kyokai on a regular basis from July 1956 and has been functioning ever since. There is an influential committee to advise the Unit in the selection of material for translation. The representation in the Committee is on a wide basis; the members are drawn from among university scholars, government officials, industrialists and businessmen. A number of translators are engaged either on a whole time basis or on a specific contract. Although progress in the beginning was somewhat slow as the organization of work necessarily took time, the Unit has been doing satisfactory work during the last several months. A large number of documents have been translated and copies of the translated material, including charts and graphs are available in the Institute's library in Calcutta and New Delhi. Several very valuable publications in the field of small-scale industries, productivity of machine industry, capital accumulation, planning, structure of wages and salaries, agricultural extension, technological coefficient etc., have been acquired and are being utilized in planning studies.

Even when the material already selected has been translated, the need for the Unit will remain, as there will be always new and interesting publications that may be expected to become available. Requirements of data change from time to time and there is a continuing need for getting comparative information from other countries, properly collected and digested under the supervision of experts, on subjects covering different fields and of interest in our planning studies.

Visit to China

The team left Japan on 8 July and reached Canton the next day. One of the Directors of the State Planning Commission, Mr. Yung Lung-Kwei had come down all the way from Peking to Canton to receive the team. For the entire period of its stay in China, Mr. Yung Lung-Kwei looked after the team and earned the admiration and gratitude of the team for his quiet efficiency, great ability and friendliness. After a stay of two days in Canton, the team left by train from Peking, halting for a few hours at Wuhan on the way. Both at Canton and Wuhan, the members of the team met officials of the planning and statistical organisations and visited institutions and sites of interest. In Peking, where the team spent most of its time, it held a series of discussions with officials of the Planning Commission, Central Statistical Board, and other Government departments and visited factories, universities, training institutions and other places of interest. As there were already at that time separate delegations from India visiting China for the study of cooperatives and agricultural planning, the team concentrated attention on the following main subjects :

(*a*) Planning machinery and organization and techniques of Planning;
(*b*) National income and its distribution;
(*c*) Scope and organization of state trading.
(*d*) Relationship between the public and the private sectors;
(*e*) Problem of crafts and attitude to cottage industry;
(*f*) Problem of unemployment;
(*g*) Policy and programme regarding heavy machine-building industry;
(*h*) Problem of organization and management of public enterprizes;

(*i*) Role of scientific research in relation to planning ;
(*j*) Structure of wages and salaries.

A detailed statement setting out the points for discussion and for collection of data had been sent in advance to the State Planning Commission in China. The Commission had taken great pains to draw up a very full programme of discussions and visits for the entire period of the team's stay in China. Discussions were held every morning from 9 a.m. to 1 p.m. in the office of the State Planning Commission and most often there was another equally long session in the afternoon. On every topic of interest to the team, as indicated in advance by the team, an expert gave a talk, at the end of which questions were invited and answers given. Services of able interpreters were available and it was possible to take down very detailed notes. All the experts were evidently very well prepared and often had written notes with them.

The visit was all too brief to enable much travel in a vast country like China. Most of the time was spent in discussions at Peking; but visits outside Peking were not altogether neglected. The team made a flying visit to the industrial areas in the North and in this visit were included Shen-yang, Anshan and Chang-chun. Apart from visiting the famous iron and steel works of Anshan, the new truck factory at Chang-chun and a number of industrial enterprises in Shen-yang, this visit gave an opportunity to the team to see the work of planning and statistical organizations at the provincial and municipal levels. Another interesting trip was to Shanghai to study the conditions of industries still in the private sector. One of the members of the team, Sri R. Natarajan, who was making a special study of industrial enterprises and labour relations, was able to visit some other industrial cities and also make a somewhat longer stay in Shanghai.

Towards the end of their stay in China, the members of the team were invited to meet Premier Chou-en-lai and dine with him on 14 August 1956 at the sea-side resort of Pei Tai Ho. The meeting lasted nearly two hours and provided the team with a unique opportunity of discussion with Premier Chou-en-lai on some of the major problems of economic development in China. The distinguished Director of the Central Statistical Board and one of the Vice-Chairmen of the State Planning Commission, Mr. Hsiu Mou Chiao and the Counseller, Indian Embassy Sri Bahadur Singh were among those who were present on the occasion.

At the request of the State Planning Commission, the members of the team delivered a few lectures at the Planning Commission premises. Pitambar Pant spoke on Planning in India. R. Natarajan talked about Industries in India and M. Mukherjee spoke on the Statistical set-up in India. M. Mukherjee gave another talk to a small group of technical workers on the method of estimation of national income in India and the logical basis of income estimation work in the West and in the East.

The team received every cooperation and assistance in its work and the warmth of friendship and the unobtrusive hospitality of the Chinese hosts left a deep impression on the members of the team.

The Indian Ambassador and the Counseller took great interest in the team's work and their advice and guidance was of great value to the team. Of great assistance to the team was the ungrudging help given by Sri Paranjape of the Embassy whose mastery of of Chinese language made it possible for the team to have a large amount of material translated from Chinese into English in record time.

The visit to China was highly educative. The team was able to get a good understanding of many aspects of development in China and was able to collect a large amount of very valuable information. The team was greatly impressed by the rapid advance in almost every field of activity in China and the purposefulness and determination, with which economic and social transformation was being attempted and being realized.

On the return of the team to India, arrangements were made to set up a small unit in Calcutta and New Delhi on a part-time basis to translate selected Chinese documents of value for planning studies. It is a matter of satisfaction that the Indian Statistical Institute has been able to develop a nucleus of a Reference and Research Unit, which will bring together, translate and analyse material of significance and value from the Chinese and Japanese sources. The existence of a very good library, copying, photostat, etc., facilities in the Institute and its wide contacts, make the task feasible.

A number of notes on various aspects have been prepared for reference. A part of the material has been put together in two reports :

1. Planning in China: Approach, Organization and Method by Pitambar Pant.
2. Industrial Planning and Growth in China by R. Natarajan.

Appendix 16: Members of the Council, 1956-57

President : Sri Chintaman D. Deshmukh.

Vice-Presidents : Dr. P. N. Banerjea, Prof. S. N. Bose, Prof. D. R. Gadgil, Sir Shri Ram.

Chairman : Sir D. N. Mitra.

Vice-Chairmen : Dr. S. K. Banerji, Sri K. P. Goenka, Sri S. C. Ray, Prof. N. R. Sen.

Treasurer : Dr. Satya Churn Law.

Secretary : Prof. P. C. Mahalanobis.

Joint Secretaries : Sri Nihar Chandra Chakravarti, Sri Pitambar Pant, Sri S. C. Sen.

Members : Srimati Chameli Bose, Prof. K. N. Chakravarti, Sri V. M. Dandekar, Sri Mohanlal Ganguly, Prof. H. Ghosh, Sri N. C. Ghosh, Dr. Q. M. Husain, Prof. D. G. Karve, Prof. K. B. Madhava, Srimati Nirmal Kumari Mahalanobis, Sri N. T. Mathew, Sri D. N. Mukherjee, Sri Moni Mohan Mukherjee, Dr. U. S. Nair, Sri R. Natarajan, Dr. B. Ramamurti, Dr. C. R. Rao, Dr. N. Sundararama Sastry, Sri Sadasiv Sengupta, Sri N. Chakravarti.

Appendix 17: Governing Body of the Research and Training School, 1956-57

Sir D. N. Mitra (*Chairman, ex-officio*), Prof. P. C. Mahalanobis (*Secretary, ex-officio*), Sri P. A. Gopalkrishnan, Sri Bali Ram Bhagat (*Representatives of the Government of India*), Dr. N. S. R. Sastry (*Reserve Bank of India*), Dr. U. Srivaraman Nair (*Inter-University Board*), Mr. George Barrell (*Associated Chamber of Commerce*), Sri D. N. Mukherjee (*Federation of Indian Chambers of Commerce & Industry*), Dr. V. G. Panse (*National Institute of Sciences*), Dr. J. P. Niyogi (*Indian Economic Association*), Sir Shri Ram, Dr. S. K. Banerjee, Prof.

S. N. Bose, Prof. K. B. Madhava, Sri N. C. Chakravarti, Dr. C. R. Rao, Sri Pitambar Pant (*Representatives of the Council of the Indian Statistical Institute*).

Finance Committee (of the Governing Body) : Sir D. N. Mitra (*Chairman, ex-officio*), Prof. P. C. Mahalanobis (*Secretary, ex-officio and Director of Research and Training School*), Sri P. A. Gopalkrishnan, Sri M. S. Bhatnagar (*Representatives of the Government of India*), Dr. N. S. R. Sastry (*Representative of the Reserve Bank of India*), Mr. George Barrell, Sri Nihar Chandra Chakravarti (*Members of the Governing Body*).

Appendix 18: Committees Set up by the Council, 1956-57

Finance Committee : Sir D. N. Mitra (*Chairman, ex-officio*), Dr. S. C. Law (*Treasurer, ex-officio*), Prof. P. C. Mahalanobis (*Hony. Secretary, ex-officio*), Dr. S. K. Banerji, Sri S. C. Ray, Sri P. A. Gopalkrishnan, Sri D. N. Mukherjee, Sri S. C. Sen, Sri N. C. Chakravarti (*Member-Secretary*).

Journal Committee : Prof. S. N. Bose, Dr. Debabrata Basu, Dr. Raghu Raj Bahadur, Prof. Panchanan Chakravarti, Dr. M. N. Ghosh, Sri D. B. Lahiri, Prof. K. B. Madhava, Prof. P. C. Mahalanobis (*Editor, Sankhyā, ex-officio*), Sri Moni Mukherjee, Dr. U. S. Nair, Dr. B. Ramamurti, Dr. C. R. Rao and Dr. N. S. R. Sastry (*Representatives of the Governing Body of the Research and Training School*), Sri A. Mahalanobis (*Member-Secretary*).

Examinations Committee : Dr. S. K. Banerjee, Sri N. C. Chakravarti, Sri V. M. Dandekar, Sri M. L. Ganguli, Sri D. Y. Lele, Prof. K. B. Mahdhava, Prof. P. C. Mahalanobis, Sri N. T. Mathew, Dr. U. S. Nair (*Representative of the Governing Body*), Sri Pitambar Pant, Dr. B. Ramamurti, Dr. C. R. Rao, Dr. N. S. R. Sastry, Sri J. M. Sen Gupta, Sri Sadasiv Sengupta, Sri K. N. Chakravarti, Sri S. R. Rao, Sri N. C. Ghosh (*Member-Secretary*).

Appendix 19: Office Bearers and Council Members of Branches, 1956-57

Bombay : Sri V. L. Mehta (*President*), Professor C. N. Vakil, Sri R. G. Saraiya, Sri L. S. Vaidyanathan and Dr. N. S. R. Sastry (*Vice-Presidents*), Sri M. A. Telang and Sri K. C. Cheriyan (*Joint Secretaries*), Dr. D. T. Lakdawalla (*Treasurer*).

Mysore : Sri A. Ananthapadmanabha Rao (*President*), Sri Srinagabhushana (*Vice-President and Secretary*), Sri R. Ramaswami (*Hony. Treasurer*), Sri Mir Ahmed Hussain (*Joint Secretary*), Sri S. K. Ekambaram, Sri R. Guru Raja Rao, Sri S. K. Rana, Sri C. Ramaswami, Sri R. Natarajan, Sri H. S. Narayana Rao and Sri S. K. Iyengar (*Members of the Executive Committee*).

Aligarh : Professor D. P. Mukherjee (*President*), Mr. A. Salim (*Secretary*).

Appendix 20: Board of Directors of the International Statistical Education Centre, Calcutta, 1956-57

Chairman : P. C. Mahalanobis.

Members : G. Darmois, R. G. D. Allen, E. Lunenberg (*International Statistical Institute*); P. A. Gopalkrishnan, S. K. Bose (alternate) (*Government of India*); C. R. Rao, S. C. Sen, A. Matthai, Secretary, (*Indian Statistical Institute*).

SANKHYĀ

THE INDIAN JOURNAL OF STATISTICS

Edited by : P. C. MAHALANOBIS

VOL. 20, PARTS 3 & 4 DECEMBER 1958

THE SCOPE OF BIOLOGICAL STATISTICS[1]

By J. B. S. HALDANE

Indian Statistical Institute, Calcutta

Statistics are always based on counting, and often on measurement. One may merely count all the individuals of a certain type, such as rice plants, in a certain area. One may count individuals of several well marked types, such as members of different castes or blood groups, fish of different species in a tank, bean pods containing zero to eight beans, and so on. Or one may measure, and then group the individuals, for example fish weighing between 1 and 1.1 kgm, or households with annual incomes between 1500 and 2000 rupees. But the fundamental data are the numbers belonging to different classes.

The individuals counted may be objects or events. Thus you may want to know how many buses there are in Calcutta, or how many passed a particular point in a given direction in every 5 minutes during 24 hours. If you do this you will count the same bus several times in the day.

A biologist may be interested in objects and events with very different scales in time and space. He may want to estimate the numbers of various kinds of cells in a cubic millimetre of human blood. This contains about 5 million red corpuscles and 10,000 white corpuscles which can be classified into about 6 different types. Or he may be concerned with the human population of India or of the world. Again he may be counting nervous impulses of which 50 or so can be detected in a nerve fibre per second, or finding from the fossil record the mean duration a genus (about 8 million years for carnivora, 40 million for mollusca). His methods will, of course, be quite different in different cases. For the small objects he will need a microscope, for the rapid events a record on a rapidly moving sensitive film. And so on. This means that a statistician must know something about methods of measurement. He must know that a beam of electrons acts as a lever without inertia in recording rapid processes, and radioactivity as a very slow clock in recording slow ones. He must know that the limits of accuracy in measuring small objects are given by wave lengths, those of visible light in microscopy, of electrons considered as wave packets in electron microscopy. If he does not understand the methods of measurement, he may grossly exaggerate the variation of what is measured. When he has data on a sample he will look for representative parameters such as means or medians, measures of spread round them, and measures of relation between different qualities or measures such as correlations.

---

[1] This is Professor J. B. S. Haldane's inaugural lecture as Research Professor in the Indian Statistical Institute. It was delivered at a public function held in the Institute on 4 October 1957 at which Professor S. N. Bose, Vice-Chancellor of Visva-Bharati presided.—*Editor.*

Let us take some examples at the microscopical level. The number of chromosomes observable in the dividing nuclei of a plant or animal is usually constant in the germ line, that is to say the cells which give rise to future generations. It is halved when the gametes, for example spermatozoa or pollen grains, are formed. The large majority of nuclear divisions forming them are regular in normal organisms. However, in a fraction of the order of one per thousand the process may go wrong, a gamete being produced with a chromosome too few or too many. But in hybrids between species these errors are much commoner, and one can sometimes classify gametes according to the number of chromosomes which they contain. As however the chromosomes may be as little as 1μ (.001 mm) in length, errors are unavoidable, and certain workers tend only to count nuclei giving a clear picture. This obviously leads to bias. Again the diameter of a human red blood corpuscle (about 7.5μ) is somewhat variable. The variance may be greatly increased in disease. But diameters are usually estimated from photographs, and we still know too little about errors of measurement. For example, Attfield (1951) in my laboratory found a mean corpuscular diameter of 7.65μ in normal, and 8.59μ in congenitally anaemic mice. But the coefficients of variation were 9.2% in the normals and 13.3% in the anaemics. The mean was increased by only 12%, the coefficient of variation by 45% or the variance by 165%. As we do not know how much of the variation is observational, the true increase is certainly greater.

The objects counted are usually individuals, but sometimes classes, sometimes both. An example of classes which are counted is given by species. For example about 200,000 species of flowering plants have been described. 21,000 of these are found in India. The only state with a larger number is Brazil. The U.K. has only about 1500, many known to be recent importations. Of course the number of species is nothing definite. But it is easier to count species than banana or banyan trees. Again if you catch N insects with a light trap they will belong to n species. The expected value of n is not far from $k \log N$, where k is a measure of the diversity of the insect population sampled. Comparative studies on these lines are needed. Little has been done in tropical countries.

The classical biometry of Karl Pearson was largely concerned with measurements such as stature, head length, strength of grip, and so on, though it must be said that Pearson's work on fertility was based on the counting of offspring. This biometrical work is certainly worthy of continuation. It is particularly important economically, since economically important biological characters, such as the milk yield of a cow, or the rice yield in grams per plant or per hectare are usually continuous variables. Before we continue I wish to consider variance as a biologist rather than a statistician. If a character has a very small coefficient of variation in a population we may expect that it is important. Deviants may be weeded out by human agency. They may be weeded out by natural selection, as for example Rendel (1943) found that very light or very heavy duck's eggs had less probability of hatching than those of median weight. Or it may be regulated physiologically as is the human temperature. An increased variance may mean a breakdown of natural or artificial selection or of physiological regulation. For example Callan and Spurway (1951) studied spermatogenesis in hybrid newts. They found that the mean number of chiasmata (junctions between homologous chromosomes) was considerably diminished. But the variance was greatly increased. That is to say the mechanism of formation of the nuclei of spermatozoa was no longer being controlled at all closely. A decrease in a correlation coefficient usually implies increased variance in an index such as a ratio of two lengths, which can often be used as a measure of shape. Thus if the length and breadth of leaves are more highly correlated on one tree than on another, this means, roughly, that their shape is more constant, or more precisely regulated.

Pearson investigated homotyposis, that is to say he discovered what fraction of the variance of a measurable or countable character, such as the number of ridges on a seed capsule, a shape parameter of a leaf, or the length of an egg, in a population, was due to variation within so called individuals. (I emphasize the unsatisfactory nature of the word individual. Palm-trees of most species seem to be individuals. They cannot be propagated by division. But most trees would more accurately be described as dividuals). This work requires extension, because, as S. K. Roy has shown, the mean of a character in different flowers of the same plant may vary during a season, as may its variance. In fact we shall have to think of each "individual" plant as defining a population of leaves, flowers, fruits, and so on, and a group of plants as a population of these populations. It might be possible to calculate generalized distances between pairs of plants, and to find the distribution of these in a large population of plants.

We may perhaps expect quicker returns from mere counting, for example counts of the numbers of seeds on different plants; and I hope that someone may do for Indian plants what Salisbury (1942) has done for British plants in this respect. One can also study natural selection, that is to say the survival and fertility of organisms differing in qualitative or measurable characters. An extremely important quantitative character is resistance to insecticides and other poisons. Workers have usually contented themselves with estimating the median lethal dose of a poison, and perhaps, though much more roughly, the dose killing 99%. It is no doubt satisfactory, for the moment, to reduce the population of mosquitoes in an area to 1% of its former value. But this involves intense selection in favour of resistance to the insecticide used, and experience has shown that in many cases the Darwinian view that such selection will increase the resistance in future generations turns out to be correct.

The classical biometricians were interested in heredity, and therefore in characters which they believed were little affected by the environment. Geneticists have naturally shown the same bias. Human blood group membership is an ideal character for geneticists, because no case is recorded with certainty in which it has changed during a human lifetime. If, however, one is interested in improving the environment one should clearly look for characters which are mainly affected by it. No measurable human characters are entirely determined by the environment, as is, for example, the language spoken at the age of five years. However, it is clear that the weight of adults, or better the weight divided by the cube of the height, is more responsive to nutritional differences than is the height. Recent American workers have chosen such measures as the thickness of a skin fold as measures of nutritional status. The urinary excretion of vitamins is another. Work on such characters will be particularly valuable in India.

Any biometrical work which does not allow of the comparison of relatives has a limited value. Though once the inheritance of such a character as blood group membership is well established, one need not make such a comparison in each population studied. The methods of non-human genetics, however, differ fundamentally from those of classical biometry in being experimental. If a geneticist suspects that a rare character is recessive he immediately mates two animals or plants bearing it, or self-fertilizes a plant. Only if all the progeny resemble their parent or parents can he conclude that it is recessive. This cannot be done with a supposedly recessive human character such as microcephalic idiocy. Animal and plant genetics almost always involve an alteration in the previous mating system of the population studied.

In consequence they involve novel statistical methods. Let me take an example. A sample of a human population includes a members of blood type M, b of type MN, and c of type N. On the null hypothesis that all types have equal survival and fertility, and that mating is at random as regards these characters, we can easily show that the expectation of the statistic $4ac-b^2+b$ is zero. But it is not quite a trivial problem to calculate its exact sampling distribution, and to establish confidence or fiducial limits. This problem is however trivial compared with that which faces an experimental geneticist. He finds D dominants and R recessives in a sample, having previously shown to his own satisfaction that the characters in question are dominant and recessive. He mates n dominants to recessives and obtains progeny. The number of progeny in the i-th of a families containing no recessives is d_i, the number in the i-th of b families containing at least one recessive is d'_i dominants and r'_i recessives. From these figures it is required to construct statistics which will enable us to test various null hypotheses, including the hypothesis that dominants and recessives are equally viable, and that mating was at random in

LIBRARY

the population sampled. I have not solved this problem, though I have given an approximate solution. I do not think that the same solution would satisfy Fisher and Neyman.

Besides changing the mating system we may try to induce mutations with radioactive substances or otherwise, and to measure the frequency with which we do so. The method which I have designed for this purpose (Haldane 1956) and which is now being used by Dr. Carter at Harwell, will, I think, measure this frequency in mice within 10% or so. Hence until a hundred induced mutations have been recorded the inaccuracy of various assumptions which I have made will not matter much. But if greater accuracy is desired, a number of extremely difficult problems will have to be solved.

A very important field of biological statistics centres round life tables. These are often based on such large numbers that sampling errors are negligible compared with other errors which beset all official statistics, and are not wholly avoided even in sample surveys. On the other hand a biologist may obtain a life table based on only a hundred animals, while being sure, either that there were no errors, or that three of the hundred were accidentally killed on known dates. The sampling theory appropriate to such data only exists in part.

Still less satisfactory are the methods for dealing with events whose probability varies with the animal's age. Within a pure line of mice, that is to say, a genetically homogeneous group of mice of the same sex, kept under fairly uniform conditions, cancer may be negligible till after the median age of death. The force of mortality from cancer, that is to say the probability that a given individual will die of cancer in a short period, then rises, and may apparently reach a more or less constant value which may be of the order of 1 to 2 percent per day. One can calculate, for a given stock under given conditions, the median age of death from cancer, that is to say the age at which half the mice would be dead if all other causes of death were eliminated. But in doing so one must make the assumption that causes of death are independent, in the sense that those mice which died, say of kidney disease, were no more and no less likely than others to die of cancer within a given time if we could have prevented the kidney disease. This is rather unlikely in a pure line, and certainly untrue in a mixed stock. Different stocks of mice differ in their susceptibility to cancer. They certainly differ in the median ages of death from cancer, defined as above. They also differ in the incidence of cancer at different sites. It is at least possible that the median ages of death from cancer at various sites are sufficient to characterize the differences between cancer incidence in different lines.

I have lately had to deal with a problem of this type in quite small groups of fish with which my wife has worked. When females are isolated, a few of them become hermaphrodites, and produce offspring without the aid of a male. And probably most of those which live for a year extrude unfertilized eggs. These events are recorded. But how are we to compare their frequencies in two groups each consisting of about sixty fish, in which the median ages of death are decidedly different?

The statistical methods which I have tried all involve dubious biological assumptions. Even if they did not, I do not see how to use the standard methods of comparison such as Student's test. Perhaps others may succeed where I have so far failed.

Any attempt to obtain life tables for a natural population of animals is extraordinarily difficult. It is possible for a natural population of plants, at least after the seeds have germinated and the cotyledons (the first leaves) are visible. But it would be extremely difficult; as in nature seedlings are commonly hidden by competing plants. The mortality must be immense, for some adult plants may produce an average of 100,000 or so seeds of which only one survives. But at least plants keep still. The best animal life table is that of Varley (1947) on a group of insects which stayed still because they spent a year in galls which they had made on plants. Varley opened about 50,000 galls, and concluded that about 1% of the larvae survived to emerge as adult flies, which then lived about a week. He enumerated and measured a dozen or so different causes of death in different parts of the life cycle.

Apart from this the best data are on birds. Fortunately most birds are fully grown before they leave the nest, so one can put numbered metal rings on their legs, and count those which are picked up after a given number of years. The fraction returned in Britain varies from under 1% for some small birds to about 15% for large ones. Most birds die in their first year of life, as they must, since each pair of small birds lay from 3 to 10 eggs, and the survivors start breeding after a year. After this a remarkably constant fraction dies each year. The fraction is about a half for birds the size of the mynah. The data so far show no signs of old age. On the other hand, data on wild sheep, which register their age by adding a new ring to their horns each year, show a great increase in mortality after about twelve years.

Sample surveys are essential when populations are very large. Let me give an example. Salt and others (1948) examined 20 cores from the same English pasture in November. They contained 42,753 arthropods, of which the majority were mites; and some were missed. There were 15,170 insects, but the numbers in a core varied from 250 to 1668, the mean being 758.5 ± 56. The density was 2.6×10^{11} per square kilometer, or 2.6×10^9 per hectare. This is the approximate number of human beings. In May the number was only 1.1×10^9 per hectare. The results are interesting as showing the extreme fluctuations in the numbers of some of the rarer groups of insects. Raw (1956) studied the Protura (primitive wingless insects) in the same area in a larger number of cores. The logarithms of $n+1$ (adding 1 because of zero counts) were fairly symmetrically distributed, and 59% lay between $\pm$ 0.3 of the mean.

The samples are usually made by mechanical methods. But biological methods can be used. Lack and Owen (*J. An. Ecol.* 25) analysed 24 meals given by swifts (an insect-eating bird) to their young, by gently pressing the throats of the babies after the mother had left them. They found an average of 680 insects per meal. These had taken about an hour to collect. The frequencies of different species were significantly different from those collected at the same time in the same area by a mechanical method, but species common in one set were usually common in the other.

In the case of plants, or of animals which do not move quickly, at least during a part of their life cycle (like birds which are very immobile as eggs) one can study their distribution in a series of unit areas within a larger unit area, for example, wire squares with sides of one metre, thrown down at random. We may find a Poisson distribution. As you know, the second factorial cumulant of a Poisson distribution is zero. If its estimate is significantly positive we may say there is a tendency to bunch. This could be due to the patchy character of the soil, the sociability of the animals in question, and so on. Where the second factorial cumulant is significantly negative we suspect some kind of antagonism between individuals or mated pairs, in some cases amounting to a territorial system. Thus in England we can divide birds into those whose nests are often close together, for example, rooks, and those where each male defends a territory round his nest, for example most song birds. In the later case the mean number of nests in a unit area is less than its variance. This is one of the simplest statistical measures of animal social behaviour.

My own work has mostly been of a rather different kind. I have worked out the consequences of various hypotheses to see how selection, inbreeding, and so on, could be expected to act on populations. In the language of probability theory I have examined a number of Markoff chains. In few cases have observational results been exact enough to confirm my conclusions more than quite roughly. However, the work was necessary in order to arrive at a logically coherent theory of evolution.

I think that the time has now come when one can see some of the observational data which are needed. An important desideratum is a careful study of the interaction between neighbouring plants. And since rice is planted out much more regularly than wheat, it offers excellent material for such work. The interaction may be highly competitive. Sakai (1957) has found that some kinds of rice greatly lower the yield of their neighbours, whether of the same race or a different one. This effect may be so great that by selecting in each generation for the highest yield one obtains plants with a lower yield than the original mixture. For highly competitive plants, though they do well when surrounded by less competitive ones, may depress one another's yield very considerably. On the other hand, Gustaffson (1953) has found, in the case of barley, that mixtures of certain races may give a higher yield than either race by itself. Such phenomena can, I think, only be studied statistically by the full use of the techniques developed by Fisher. Their explanation will call for biochemical rather than statistical research. But the facts still demand much further statistical investigation. Neither Sakai nor Gustaffson has made all the experiments possible with their own material.

The improvement of animal breeds is based on artificial selection. This is a relatively simple matter in the case of sheep where both sexes produce wool, and males can be selected for their performance. It is much more difficult in cattle bred for milk production or poultry for egg production, where only the performance of females can be measured, and yet improvement is mainly by selection of males, since only a small fraction of males is bred from. Ideally one would like to choose a bull who has begotten say ten daughters whose mean milk yield was notably above the mean of their mothers' yield. To do so one might have to wait till the bull was ten years old. This is practicable when artificial insemination is used. But in ordinary breeding one must be content with choosing a bull on the record of his mother, his grandmothers, his aunts, and so on, and perhaps some sisters, or half sisters. Dickerson and Hazel (1944)) examined this problem statistically, and concluded that progress is quicker if the bulls are merely chosen on this basis and not on the performance of their daughters. If one waits for progeny tests the steps in improvement are larger, but there are fewer of them in twenty years. This question can only be answered by statistical methods, and the answer given for cattle is not necessarily valid for goats or poultry.

The improvement of plants is a more complicated matter, as will be seen from a consideration of some of the commonest trees round the Indian Statistical Institute. Bananas (*Musa* spp.) are asexually propagated. The differences between members of the asexual progeny of a single plant are usually slight and rarely inherited. The bananas systematically grown in Bengal belong to quite a small number of clones.[1] It should be possible to determine the mean yields of these clones as a function of soil conditions, time of year, and so on, and plan planting accordingly. There is no reason to think that the mean yields of the different clones have anything like a normal distribution. So the analysis of variance into within-clone and between-clone components might be somewhat misleading. The mango is also asexually propagated, although grafting slightly complicates the problem. It is possible to make new clones, both of the banana and mango. But they cannot be produced by simple selection.

On the other hand the palms are propagated by sexual reproduction like most animals. In some species such as the *khajur*, *Phoenix sylvestris*, the sexes are separate. But so far as I know it has not even been determined which sex yields most *gur* or toddy. The coconut palm or *narikel*, *Cocos nucifera*, is a hermaphrodite like most plants. It could and should be improved. But little can be done until the father as well as the mother of each nut planted to produce a new tree is known. This knowledge can only be obtained by climbing the trees, first to protect the young female flowers from pollen, and then to pollinate them from a known tree as father. Only then will it be possible to discover the effects of self-fertilization and the regression of a tree's yield on those of its parents.

I mention these facts simply to emphasize that even where, as with the palms, no genetics at all are known, the statistician will be well advised to learn some elementary biology.

I will indicate a few further fields of biology where I think statistical methods are needed. They have already been applied to some extent to the study of animal behaviour, but mainly, if I may say so, at a rather low statistical level. Some behaviour consists in the very frequent repetition of the same act, or pair of contrary acts, for example of a fish rising for a breath of air, and a sitting bird flying away from its eggs and back to them. It has been usual only to record the frequency of the act, or the mean interval between acts, the variance of this interval being merely used to test the significance of the difference between means recorded under different circumstances. However other moments of the interval distribution are just as interesting. Moreau (1939) found that intervals of absence from a nest had a smaller coefficient of variation than intervals sitting on eggs. This is intelligible, as too long an absence could cool the eggs, perhaps dangerously. Spurway and I found that the intervals between successive ascents for air of *Anabas testudineus* had a nearly normal distribution with a coefficient of variation about 25%. By altering the gas breathed, we could lengthen or shorten the mean interval, leaving the coefficient of variation and the shape of the distribution unaltered. Or we could barely alter the mean, but increase the variance about forty times, and make the distribution highly skew positively.

---

[1] In this context a clone means all the plants derived from a single seedling by asexual propagation.

The application of statistics to experimental animal psychology has largely been to learning. I believe that it will be much more fruitful, at least for the next few years, to study activities which are either unlearned or completely learned, so that the learning process does not alter the animal's behaviour during our observation. I should like to study animal technology, and in particular the construction of nests by birds and webs by spiders. Consider, for example, a weaver-bird's nest. It may be made mainly from one species of grass. What is the mean number of blades used, and its variance between nests ? What is the mean length of blade used in a given nest, and its variance ? Do these vary between nests ? The results could be compared with data on human craftsmen.

I have said nothing about the application of statistics to human psychology. This seems to have developed in rather serious isolation from other statistics, and it is at least possible that a wholly fresh attack on it might be useful. Nor have I dealt with human medical statistics. The mathematical principles involved are seldom difficult. On the other hand, the data are biassed for a variety of reasons. For example, the fraction of cases of a disease reported varies with class and occupation. This is obviously true in India, but it is also true in Britain. If coalminers lose more days through sickness than clerks, this is not wholly because coalmining is the less healthy occupation. It is partly because coalminers are incapacitated by an illness which would not incapacitate a clerk. Again old coalminers have a rather surprisingly low mortality. This is partly because those who suffer from illness in middle age leave the occupation and are registered as shop assistants, caretakers, or members of some other not too strenuous occupation, when they die.

Until the number of practising physicians in India is increased at least five times it is hopeless to expect adequate medical statistics in India. It is more important to save lives than to keep detailed records of death and illness. And there are not enough men and women to do both. It is not impossible that valuable information might be gained by Sample Surveys, but their planning would demand much knowledge of social and economic facts as well as of statistical theory. I have only been able to help medical statisticians when they were very sure of their data.

In the diagnosis of disease, especially chronic disease, physicians are apt to rely on deviations from the median in one or more quantitative physiological characters. The mouth temperature and blood glucose concentration are well-known examples. However, the distributions of some such measurable characters are likely to be rather different in India and Europe, certainly on account of differences in climate and diet, and possibly on account of racial differences. For example, on going to a hot climate, Europeans usually increase their blood volume by 10% or so, thus diluting their haemoglobin. The blood has many functions. One is cooling the body by moving quickly through the skin. For this function a high volume and low viscosity are useful. I should thus expect healthy people in India to have a rather larger blood volume, with less haemoglobin and fewer corpuscles per millilitre, than comparable healthy Europeans. If so the criterion for a diagnosis of anaemia should be more strict than in Europe. We may hope that the research organized by Dr. B. C. Das in this Institute will furnish data on a number of measurable physiological and biochemical characters which will provide better norms for Indians than exist at present.

Statistical palaeontology is becoming important for two very different reasons. On the one hand, it gives us almost the only method of measuring rates of evolution directly. I can perhaps claim to have opened this particular field (Haldane 1944) by obtaining very rough measures of the rates at which the means of measurable characters changed in well established evolutionary lines such as the ancestors of horses. Mean rates were of the order of $3\frac{1}{2}\%$ per million years. On the other hand the means of certain characters in fossil molluscs are used to characterize particular British coal seams. The materials available are, however, biassed in several ways. Smaller and thinner bones and shells are more likely to be destroyed, or to be thrown away by collectors, who will on the other hand pick out grossly abnormal specimens even if they are damaged. A statistician who wishes to work on such data should have at least some experience of field work in palaeontology.

I have said little of the mathematical problems raised. Pearson's and Fisher's researches both arose from the consideration of biological questions. But they have not exhausted the mathematical side of the subject. The question of experimental design has only been solved by Fisher for an important but limited class of experiments. Fisher's technique is based on the fact that agricultural soil is very variable. It is possible to standardize soil and other conditions in a green-house. Went (1953) has reduced the coefficient of variation of the height of individual plants in a pure line to 4%. Where such standardization is possible the question of experimental design is essentially one of costing. Standardization is expensive, but if it reduces the number of plants needed by 95%, it may be worth while, provided we are answering a biological question, and not trying to answer an economic one. For cultivators are certainly not going to keep rice plants in individual pots with controlled humidity, lighting and so forth.

The work of Bonnier and his colleagues (e.g. 1948) in Sweden gives a beautiful example of this kind of planning. He wished to work on the influence of diet, in childhood and later, on the milk yield of cows. Is it economical to feed them very well as calves, or can they be left to find their own food, and only pampered during pregnancy and lactation? He found that the mean square difference between milk yields of pairs of monozygotic twins was about five percent of that between calves taken at random from the same herds. He therefore collected such twin pairs from all over Sweden, at considerable expense, and brought them up side by side in his cattle sheds on different diets. This reduced his expenditure on food, shelter, and attendance by about 95%, which far more than balanced the cost of advertising, transport, and tests on blood and other characters of the twin pairs which were needed to establish that they were monozygotic. It should never be forgotten that the planning of biological experiments, like that of sample surveys, is an economic matter.

An unsolved mathematical problem is the statistical treatment of growth curves. Human male heights, for example, increase rather suddenly during a year or two at the age of puberty. But this age is variable. So the graph of the average height of 100 boys shows a slow rise spread over many years, and is not typical of any individual. The problem of finding a representative growth curve, and estimating the variation round it, is difficult. It would be interesting, however, to collect even twenty such curves in India for comparison with European and American data.

I am at present engaged in working out the theory of estimating rates of increase of populations with known birth rates. The rate at which a population will increase exponentially depends mainly on two parameters, the mean number of daughters per mother, and the mean maternal age at the birth of a daughter. The other moments of the maternal age distribution are less important, though they cannot be neglected. However, all these statistics are correlated when individual mothers are considered, and it is hard to give even the approximate sampling variance of the rate estimate, and would be very hard indeed to give its exact distribution, since the statistics for individual mothers do not have normal distributions.

I hope I have shown you that the field of biological statistics is wide open. Professor Rao has improved and standardized methods in what may be called the field of classical or Pearsonian biometry. But when we leave such data as measurements on adult human beings, distributions may be very far from normal. Salisbury (1942) gives figures for the numbers of seeds or fruits borne by different plants which are about as positively skew as the distribution of real incomes in India. That is to say, even the distribution of their logarithms is still positively skew. For nobody could live on one tenth of the geometrical mean income, and plenty of people have more than ten times this income. It is a striking fact that if two or more variates are fairly highly correlated, for example length, breadth and depth in fish of the same species but different sizes, one can apply much correlation technique even if all the distributions are extremely skew. For the breadths and depths of fish of a given length may have fairly normal distributions. Such methods are likely to be particularly useful in statistical palaeontology.

I hope I have shown you that the field of biological statistics is very broad, and that there is still plenty of work to be done in it, both practical and theoretical. I would only warn you that practical work will probably be inefficient without some theoretical knowledge, and theoretical work may be ornamental but will not be useful without some practical experience. It should be one of the tasks of this Institute to see that theory and practice are not too widely separated.

References

Attfield, M. (1951) : Inherited macrocytic anaemias in the house mouse. III. Blood cell diameters. *J. Genet.*, **50**, 250.

Bonnier, G., Hansson, A., and Skjervold, H. (1948) : Studies on monozygous cattle twins. IX. The interplay of heredity and environment on growth and yield. *Acta. Agric. Suec.*, **3**, 1.

Callan, G., and Spurway, H. (1951) : A study of meiosis in interracial hybrids of the newt, *Triturus cristatus*. *J. Genet.*, **50**, 235.

Dickerson, G. E., and Hazel, L. N. (1944) : The effectiveness of selection on progeny performance as a supplement to earlier culling in livestock. *J. Agric. Res.*, **69**, 459.

Gustaffson, A. (1953) : The cooperation of genotypes in barley. *Hereditas*, **39**, 1–18.

Haldane, J. B. S. (1949): Suggestions as to quantitative measurement of rates of evolution. *Evolution*, **3**, 51.

——— (1956) : The detection of autosomal lethals in mice induced by mutagenic agents. *J. Genet.*, **54**, 327.

Lack, D., and Owen, B. F. (1955) : The food of the Swift. *J. An. Ecol.*, **24**, 120–136.

Moreau, R. E. (1939) : Numerical data on African birds' behaviour at the nest : *Hirundo s. Smithii* Leach, the Wiretailed Swallow. *Proc. Zool. Soc.*, **109**A, 109–125.

Raw, F. (1956) : The abundance and distribution of protura in grassland. *J. An. Ecol.*, **25**, 15–21.

Rendel, J. M. (1943) : Variations in the weights of hatched and unhatched ducks' eggs. *Biometrika*, **33**, 48.

Salisbury, E. J. (1942) : *The Reproductive Capacity of Plants.* London, Bell.

Salt, G., Hollick, F. S. I., Raw, F., and Brian, M. V. (1948) : The arthropod population of pasture soil. *J. An. Ecol.*, **17**, 139–150.

Sakai, K. (1957) : Studies on competition in plants. VII. Effect on competition of varying numbers of competing and non-competing individuals. *J. Genet.*, **55**, 227.

Varley, G. C. (1947) : Natural control of population balance in the knapweed gallfly (*Urophora jaceana*). *J. An. Ecol.*, **16**, 139.

Went, F. W. (1953) : Gene action in relation to growth and development. I. Phenotypic variability. *Proc. Nat. Ac. Sci.*, Washington, **39**, 839.

Paper received : July, 1958.

EXAMPLES OF INCONSISTENCY OF MAXIMUM LIKELIHOOD ESTIMATES

By R. R. BAHADUR

Indian Statistical Institute, Calcutta

SUMMARY. This note contains two examples, concerning independent observations from a fixed population of real values, where maximum likelihood estimates of the population distribution function do not converge, as the sample size tends to infinity, to the actual distribution function. Such examples are of interest since in previously published examples of the failure of the method, the estimated distribution, if it exists at all, does converge to the actual one.

The maximum likelihood (m.l.) method estimates the entire population distribution from given data, even if the statistician is interested only in some particular parameter, i.e., some functional of the population distribution. It therefore seems interesting and appropriate, if not logically necessary, to enquire whether the method is consistent in the sense that, in increasing samples of independent observations from a given population, the estimated distribution converges, in a given intrinsic sense, to the actual one. According to this viewpoint, the consistency of the m.l. estimate of a particular parameter or set of parameters (e.g., a set specifying the population distribution) is a subsidiary problem, determined mainly by such non-stochastic questions as whether the parameters of interest are identifiable or whether they are continuous functionals of the population distribution.

Wald's famous proof of the consistency of m.l. estimates (Wald, 1949; c.f. also Kiefer and Wolfowitz, 1956) can easily be formulated so as to yield regularity conditions sufficient for consistency in the sense described above. However, the need for regularity conditions, except in so far as they guarantee the existence of m.l. estimates, is not clear from the examples in the literature (Basu, 1955; Kiefer and Wolfowitz, 1956; Lecam and Kraft, 1956). In some of these examples, the likelihood function is unbounded, so that m.l. or even approximate m.l. estimates do not exist. In the others, where m.l. or approximate m.l. estimates do exist, it can be shown that the estimated distribution converges to the actual distribution in the strong sense that, with probability one, the estimated probability of any event in the sample space of a single observation tends, uniformly in events, to the correct probability. It seems to the writer that in such examples the m.l. estimate can hardly be claimed to be inconsistent, since the claim must be based on convergence definitions remote from statistics, or on the failure of the estimated value of a discontinuous functional to converge to the population value.

The two examples that follow exhibit m.l. estimates that fail to converge to the population distribution even in the weakest of intrinsic senses. In both examples, a single observation from the population is denoted by X, and X takes real values. For definiteness, it is understood that a sequence P_1, P_2, ... of probability distributions of X 'converges' to a probability distribution Q if $P_k\ (X \leqslant t) \to Q(X \leqslant t)$ at every continuity point of the latter distribution function as $k \to \infty$.

Example 1. (*The continuous case*). Suppose that it is known only that X is distributed according to some Q such that

$$dQ = f(x)dx, \ \int_0^1 f(x)dx = 1, \qquad \dots \quad (1)$$

where f is continuous and

$$0 \leqslant f(x) \leqslant 2 \qquad \dots \quad (2)$$

for $0 \leqslant x \leqslant 1$. Given n independent observations on X, say $X_1, X_2, \dots$ and X_n, it is clear that P_n is an m.l. estimate of Q if and only if $dP_n = f_n(x)dx$, where f_n is continuous and satisfies (1), (2) and

$$f_n(X_i) = 2 \text{ for } i = 1, 2, \dots, n. \qquad \dots \quad (3)$$

The inconsequential requirement (3) is satisfied by many inconsistent estimates and also by some consistent ones. The verification is omitted. It might appear at first sight that m.l. estimation fails here because the set of alternative distributions of X is infinite dimensional, but it is easy to construct versions of the same example in which the set in question is only countably infinite.

In the preceding example, the m.l. principle provides no indication of the actual distribution. In the following one, the principle provides a definite but ultimately misleading indication.

Example 2. (*The discrete case*). In this example, (i) X is confined to the non-negative integers $j = 0, 1, 2, \dots$ ad inf.; (ii) the set of alternative distributions of X is a countably infinite set $\{P_1, P_2, \dots, P_\infty\}$, where

$$\lim_{k \to \infty} P_k = P_\infty; \qquad \dots \quad (4)$$

(iii) for each $n = 1, 2, \dots$ and any set $X_1, X_2, \dots, X_n$ of n sample values, the (or an) m.l. estimate exists; and (iv) no matter what the actual distribution of X, *any* sequence of m.l. estimated distributions converges to P_∞; more precisely, if $h_n = h_n(X_1, X_2, \dots, X_n)$ is the least index h such that P_h is an m.l. estimate based on the first n observations, then

$$P_k^{(\infty)} \, [\lim_{n \to \infty} h_n = \infty] = 1 \qquad \dots \quad (5)$$

for each $k = 1, 2, \dots, \infty$, where $P_k^{(\infty)}$ denotes the probability measure on the space of infinite sequences $X_1, X_2, \dots$ when X is distributed according to P_k.

Let

$$\alpha_j = 1 + 2 \left[\frac{\log \log \log (j+a)}{\log \log (j+a)} \right] \qquad \dots \quad (6)$$

for $j = 1, 2, \dots$ where a is a constant such that α_j is well-defined and positive for $j \geqslant 1$, e.g. $a = 27$. Next, let

$$m_j = b/(j+a)[\log (j+a)]^{\alpha_j} \qquad \dots \quad (7)$$

for $j = 1, 2, \dots$ where b is chosen so that

$$m_j > 0, \ \sum_{j=1}^{\infty} m_j = \frac{1}{4}. \qquad \dots \quad (8)$$

Define

$$P_{\infty}(X = j) = \begin{cases} 3/4 & \text{for } j = 0 \\ m_j & \text{for } j > 0 \end{cases} \qquad \dots \quad (9)$$

and for each $k = 1, 2, \dots$ define

$$P_k(X = j) = \begin{cases} 3/4 - d_k & \text{for } j = 0 \\ m_j & \text{for } j > 0, j \neq k \\ \sqrt{m_k} & \text{for } j = k \end{cases} \qquad \dots \quad (10)$$

where

$$d_k = \sqrt{m_k} - m_k. \qquad \dots \quad (11)$$

Since $0 < d_k < \frac{1}{2}$ and $d_k \to 0$ as $k \to \infty$ by (8), it is clear from (8), (9), (10) and (11) that $P_1, P_2 \dots, P_{\infty}$ is a set of distinct probability distributions, and that (4) holds.

Now let $X_1, X_2, \dots$ be a sequence of independent and identically distributed observations on X. For any $n = 1, 2, \dots$ let $f_{jn} = f_{jn}(X_1, \dots, X_n)$ denote the frequency of the event $\{X = j\}$ in the first n observations and let $L_n(k) = L_n(X_1, X_2 \dots, X_n \mid k)$ be the logarithm of the likelihood function. Then, for $k < \infty$, we have

$$L_n(k) - L_n(\infty) = f_{0n} \log [1 - (4d_k/3)] + (\tfrac{1}{2}) f_{kn} \log (1/m_k), \qquad \dots \quad (12)$$

by (9) and (10). Let

$$Y_n = \max\{X_1, X_2, \dots, X_n\}. \qquad \dots \quad (13)$$

Since $f_{jn} = 0$ for all $j > Y_n$, it follows from (12) that $L_n(k) \leqslant L_n(\infty)$ for all $k > Y_n$. Consequently there exists an h such that $L_n(h) = \sup_k\{L_n(k)\}$ (i.e. an m.l. estimate exists), and there must be such an h in the set $\{1, 2, \dots, Y_n, \infty\}$. Let h_n be the smallest maximizing h. We proceed to show that (5) holds. It will be shown, incidentally, that

$$P_k^{(\infty)} [h_n \leqslant Y_n \text{ for all sufficiently large } n] = 1 \qquad \dots \quad (14)$$

for each $k = 1, 2, \dots, \infty$. It is interesting to note that (14) holds even if h_n is the *largest* maximizing index, so that the distribution P_{∞}, to which all m.l. estimates converge, is comparatively unattractive to the m.l. principle, even when P_{∞} itself obtains.

Choose and fix a $k = 1, 2, \dots,$ or ∞ and assume henceforth that X is distributed according to P_k. Since $Y_n \to \infty$ with probability 1, since $f_{jn} \geqslant 1$ for $j = Y_n$, and since

$$\log \left(\frac{1}{m_j} \right) \geqslant \log (j+a) - \log (b) \qquad \dots \quad (15)$$

for $j \geqslant 1$ by (6) and (7), it follows from (12) that

$$\liminf_{n\to\infty} \left\{ \frac{L_n(Y_n)-L_n(\infty)}{n} \right\} \geqslant \liminf_{n\to\infty} \left\{ \frac{\log Y_n}{2n} \right\} \qquad \ldots \quad (16)$$

with probability one. Now let M be an arbitrary positive integer. It follows from (12) that

$$\max_{1 \leqslant j \leqslant M} \left\{ \frac{L_n(j)-L_n(\infty)}{n} \right\} \leqslant N \qquad \ldots \quad (17)$$

where N is a constant depending only on M. Suppose for the moment that

$$P_k^{(\infty)} \left[\lim_{n\to\infty} \left\{ \frac{\log Y_n}{n} \right\} = \infty \right] = 1. \qquad \ldots \quad (18)$$

It will then follow from (16), (17) and the definition of h_n that, with probability one, $M < h_n \leqslant Y_n$ for all sufficiently large n. Since M is arbitrary, this will establish (5) and (14).

It remains therefore to verify (18). For any $z \geqslant 1$, let m_z be defined by writing z instead of j in (6) and (7). Then m is a decreasing function of z. Hence we have

$$P_k(X \geqslant z) = \sum_{j \geqslant z} m_j \geqslant \int_{z+1}^{\infty} m_z \, dz = \frac{b}{\log \log(a+z+1)} \qquad \ldots \quad (19)$$

for all $z \geqslant 1$ if $k = \infty$ and for all $z > k$ if $k < \infty$, by (9) and (10). Let c be a constant, $0 < c < \infty$. A straightforward computation using (19) shows that

$$\sum_{n=1}^{\infty} P_k^{(\infty)} [\log Y_n < cn] = \sum_{n=1}^{\infty} [P_k(X < e^{cn})]^n < \infty. \qquad \ldots \quad (20)$$

Hence, by the Borel-Cantelli lemma, $\log Y_n \geqslant cn$ for all sufficiently large n, with probability one. Since c is arbitrary, (18) is established, and this completes the verification of the example.

In conclusion, the writer wishes to thank his associates D. Basu, C. R. Rao, and V. S. Varadarajan for several discussions and helpful suggestions.

References

Basu, D. (1955): An inconsistency of the method of maximum likelihood. *Ann. Math. Stat.*, **26**, 144.

Kiefer, J. and Wolfowitz, J. (1956): Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, **27**, 887.

Kraft, C. and LeCam, L. (1956): A remark on the roots of the maximum likelihood equation. *Ann. Math. Stat.*, **27**, 1174.

Wald, A. (1949): Note on the consistency of the maximum likelihood estimate. *Ann. Math. Stat.*, **20**, 595.

Paper received : February, 1958.

MAXIMUM LIKELIHOOD ESTIMATION FOR THE MULTINOMIAL DISTRIBUTION WITH INFINITE NUMBER OF CELLS

By C. RADHAKRISHNA RAO

Indian Statistical Institute, Calcutta

SUMMARY. Maximum likelihood (m.l.) estimate of the infinite multinomial distribution exists with probability 1 and is consistent under a simple condition on the cell probabilities. To prove the consistency of an m.l. estimate of a parameter it is necessary to assume that the parameter is a continuous function of the distribution. The existence of the m.l. estimates of parameters and their consistency is established under conditions slightly weaker than those assumed by earlier writers.

1. INTRODUCTION

The object of this paper is to extend the results obtained earlier (Rao, 1957) in the problem of maximum likelihood estimation for the finite multinomial distribution (f.m.d.) to the case of a multinomial distribution with infinite number of cells, which may be denoted by (i.m.d.).

As in the case of the f.m.d., the consistency of the maximum likelihood (m.l.) estimate of the i.m.d. is established first, and then the properties of m.l. equation estimates are studied.

In the parametric case, it has been assumed for simplicity in presentation that there is only one unknown parameter. But the method of proof given here holds good for any number of unknown parameters without any modification or without any extra results being proved.

2. DEFINITIONS AND PRELIMINARY LEMMAS

The hypothetical frequencies in the infinite number of classes are represented by $(\pi_1, \pi_2, \ldots)$ and the observed relative frequencies by $(p_1, p_2, \ldots)$. All summations in the following definitions and in the rest of the paper are taken from 1 to ∞ unless otherwise stated.

Definition 1 : The m.l. estimate of the i.m.d. is a distribution $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \ldots)$, if it exists, belonging to a given admissible class A of distributions π, for which the expression

$$\Sigma\ p_i \log \pi_i \text{ is a maximum} \qquad \ldots \quad (2.1)$$

when π is restricted to A.

Definition 1.2 : π^* is said to be an approximate or near m.l. estimate if

$$\Sigma\; p_i \;\log\; \pi_i^* \geqslant \frac{\log c}{n} + \sup_{\pi \,\epsilon\, A}\; \Sigma\, p_i \;\log\; \pi_i \qquad \dots \quad (2.2)$$

where n is the sample size and $0 < c < 1$.

Definition 2 : The m.l. estimate of θ occurring in the specification of the hypothetical frequencies is a value of θ, if one exists in the admissible set of values of θ, for which

$$\Sigma\, p_i \;\log\; \pi_i\,(\theta) \text{ is a maximum.} \qquad \dots \quad (2.3)$$

Near m.l. estimate of θ can be defined in the same way as in definition (2.2).

Definition 3 : The m.l. equation is

$$\Sigma\; \frac{p_i}{\pi_i}\,\frac{d\pi_i}{d\theta} = 0. \qquad \dots \quad (2.4)$$

Definition 4 : The maximum likelihood equation (or m.l.e.) estimate is that root (or a root) of the likelihood equation which provides the maximum of the likelihood when θ is restricted to the roots.

The m.l.e. estimate is not defined if the equation (2.4) has no roots.

In what follows, we shall denote a sequence of independent observations from the given multinomial population by $X_\infty = (x_1, x_2, \dots)$.

Definition 5 : A sequence of functions $t_n = t_n(x_1, \dots, x_n)$, $n = 1, 2, \dots$ ad inf, is said to be a *consistent root of the likelihood equation* if, except for a set of sequences $X_\infty = (x_1, x_2, \dots$ ad inf) of probability zero,

(i) t_n is well-defined for all sufficiently large n,

(ii) t_n is a root of the likelihood equation for all sufficiently large n, and

(iii) $t_n \rightarrow \theta_0$ as $n \rightarrow \infty$, where θ_0 is the actual parameter value.

In Definition 5, the phrase 'all sufficiently large n' means that for any given sequence X_∞ there is an m, depending on the sequence, such that the statement is true for all $n > m$.

Definition 6 : For each β in some index set let $\{t_{n\beta}\}$ be a sequence of estimates of θ. This family is said to be uniformly consistent if

$$u_n = \inf_{\beta}\; \{t_{n\beta}\} \text{ and } v_n = \sup_{\beta}\; \{t_{n\beta}\}$$

are themselves consistent estimates of θ.

Lemma 1 : *If* $\Sigma\, a_i$ *and* $\Sigma\, b_i$ *are two convergent sequences of non-negative numbers such that* $\Sigma\, a_i \geqslant \Sigma\, b_i$, *then*

$$\Sigma'\, a_i \log \frac{b_i}{a_i} \leqslant 0, \qquad \ldots \quad (2.5)$$

where the summation Σ' *is extended over non-zero values of* a_i. *The equality is attained when and only when* $a_i = b_i$ *for all* i.

The proof of Lemma 1 is similar to that given by Rao (1957) when the number of elements is finite, following a method used by Kullback and Liebler (1951).

Lemma 2 : *If the hypotheses of Lemma* 1 *are satisfied and if* $a_i \leqslant 1$, $b_i \leqslant 1$ *for all* i, *then*

$$2\, \Sigma'\, a_i \log \frac{a_i}{b_i} \geqslant \Sigma'\, a_i(a_i - b_i)^2. \qquad \ldots \quad (2.6)$$

To establish this, we note that for $x > 0$,

$$\log\, x = (x-1) - (x-1)^2\, \frac{1}{2y^2} \text{ with } y \in (1, x).$$

It follows hence that

$$\Sigma'\, a_i \log \frac{b_i}{a_i} = \Sigma'\, b_i - \Sigma'\, a_i - \Sigma' a_i\, \frac{(b_i - a_i)^2}{2(z_i^2)} \qquad \ldots \quad (2.7)$$

where $z_i^2 \in (a_i^2, b_i^2)$ and hence $\leqslant 1$ for each i. Changing the signs on both sides of (2.7) and observing that $\Sigma'\, a_i - \Sigma' b_i$ is not less than zero, we obtain

$$2\, \Sigma'\, a_i \log \frac{a_i}{b_i} \geqslant \Sigma'\, a_i\, \frac{(b_i - a_i)^2}{z_i^2} \geqslant \Sigma'\, a_i(b_i - a_i)^2$$

as desired.

3. Consistency of M.L. estimates

Let A be the given admissible set of i.m. distributions, a typical member of which is denoted by $\pi = (\pi_1, \pi_2, \ldots)$. We shall make the following assumption about the true distribution $\pi^0 = (\pi_1^0, \pi_2^0, \ldots)$:

Assumption A_1 : $\Sigma\, \pi_i^0 \log \pi_i^0 > -\infty$.

The consistency of the m.l. estimate of the i.m.d. is established under this sole assumption. The consistency of m.l. estimates under the same assumption can be also proved by using the arguments of Wald (1949), as pointed out by Kiefer and Wolfowitz (1956). The proof given here, however, seems simpler and more direct, It may be worthwhile to note here that Assumption A_1 is not necessary for consistency. For, if we allow A to be the set of all possible i.m. distributions, the assumption A_1 is clearly not satisfied but the m.l. estimate, which is then the observed distribution, is consistent. On the other hand, it is known that m.l. estimates can be inconsistent when A_1 is not true (Bahadur, 1958).

First let us suppose that the m.l. estimate exists and let us represent it by $\hat{\pi} = (\hat{\pi}_1, \hat{\pi}_2, \ldots)$. If $p = (p_1, p_2, \ldots)$ represents the observed cell proportions, then

$$\Sigma\, p_i \log \pi_i^0 \leqslant \Sigma\, p_i \log \hat{\pi}_i \leqslant \Sigma\, p_i \log p_i. \qquad \ldots \quad (3.1)$$

By the strong law of large numbers

$$\Sigma\, p_i \log \pi_i^0 \rightarrow \Sigma\, \pi_i^0 \log \pi_i^0; \qquad \ldots \quad (3.2)$$

where in what follows the symbol→, when applied to random variables, indicates convergence with probability 1. For any given k and any sample sequence $X_\infty = (x_1, x_2, \ldots, x_n, \ldots)$ we have

$$\Sigma\, p_i \log p_i \leqslant \sum_1^k p_i \log p_i$$

for every n, and therefore

$$\limsup_{n \to \infty} \Sigma\, p_i \log p_i \leqslant \limsup_{n \to \infty} \sum_1^k p_i \log p_i.$$

Consequently, $\limsup\limits_{n \to \infty} \sum\limits_{t} p_i \log p_i \leqslant \sum\limits_1^k \pi_i^0 \log \pi_i^0$ with probability 1. Since k is arbitrary, it follows that

$$\limsup_{n \to \infty} \Sigma\, p_i \log p_i \leqslant \Sigma\, \pi_i^0 \log \pi_i^0 \qquad \ldots \quad (3.3)$$

with probability 1.

On the other hand, $\Sigma\, p_i \log p_i \geqslant \Sigma\, p_i \log \pi_i^0$ by (3.1). Hence

$$\liminf_{n \to \infty} \Sigma\, p_i \log p_i \geqslant \liminf_{n \to \infty} \Sigma\, p_i \log \pi_i^0 = \Sigma \pi_i^0 \log \pi_i^0 \text{ with probability 1.} \quad \ldots \; (3.4)$$

From (3.3) and (3.4) it follows that

$$\Sigma\, p_i \log p_i \rightarrow \Sigma\, \pi_i^0 \log \pi_i^0$$

and using (3.2) we see that all the three terms in (3.1) tend to the same limit. Since this common limit is finite by assumption A_1, we have

$$\Sigma\, p_i \log \frac{\hat{\pi}}{p_i} \rightarrow 0.$$

Hence, by an application of Lemma 2,

$$\Sigma\, p_i\, (\hat{\pi}_i - p_i)^2 \rightarrow 0,$$

hence $p_i\, (\hat{\pi}_i - p_i)^2 \rightarrow 0$ for each i. This implies that $(\hat{\pi}_i - \pi_i^0) \rightarrow 0$ for each i for which $\pi_i^0 \neq 0$. Consequently, $\hat{\pi}_i \rightarrow \pi_i^0$ for every i. Thus the m.l. estimate of each of the individual cell proportions tends to the true value. According to a theorem of Scheffe (1947), this is equivalent to

$$\Sigma \mid \pi_i - \pi_i^0 \mid \rightarrow 0 \qquad \dots \quad (3.5)$$

which exhibits the cônsistency of the m.l. estimate of the entire distribution in a strong form.

When the m.l. estimate does not exist but near m.l. estimates are considered, their consistency can be proved by a modification of the preceding argument (cf. Rao, 1957). The modified argument is presented in Theorem 1 below for the parametric case.

As observed earlier (Rao, 1957), the estimation of a parameter presents some difficulty unless it is suitably related to the admissible distributions. If $A = \{\pi\,(\theta)\}$ is a given parametric representation of the set A of admissible distributions, with θ a real parameter, and θ_0 is the true value of the parameter, and assumption A_1 holds, then the consistency of the m.l. estimate of the parameter follows immediately, provided only that the following continuity assumption holds:

Assumption A_2 : If $\{\theta_m\}$ is a sequence of parameter values for $m = 0, 1, 2, \dots$ such that for each $i = 1, 2, \dots, \pi_i(\theta_m) \rightarrow \pi_i(\theta_0)$ as $m \rightarrow \infty$, then $\theta_m \rightarrow \theta_0$ as $m \rightarrow \infty$.

We shall now prove the following theorem.

Theorem 1: *If* A_1 *and* A_2 *hold, then the class of all approximate m.l. estimates (definitions 1.2 and 2) corresponding to a fixed* c, $0 < c \leqslant 1$, *is uniformly consistent (definition 6).*

To prove this result, we observe first that if $\theta^* = \theta_n(x_1, x_2, \dots, x_n)$ is *any* approximate m.l. estimate we have

$$\begin{aligned} \Sigma\, p_i \log \pi_i(\theta_0) &\leqslant \sup_{\theta}\, \Sigma\, p_i \log \pi_i(\theta) \\ &\leqslant -\frac{\log c}{n} + \Sigma p_i \log \pi_i(\theta^*) \\ &\leqslant -\frac{\log c}{n} + \Sigma p_i \log p_i. \end{aligned}$$

Hence

$$\Sigma\, p_i \log \frac{\pi_i(\theta_0)}{p_i} + \frac{\log c}{n} \leqslant \Sigma\, p_i \log \frac{\pi_i(\theta^*)}{p_i} \leqslant 0. \qquad \ldots \quad (3.6)$$

It has already been shown above that under A_1, $\Sigma\, p_i \log \frac{\pi_i(\theta_0)}{p_i} \to 0$. It follows hence by (3.6) that

$$\inf_{\theta^*} \Sigma\, p_i \log \frac{\pi_i(\theta^*)}{p_i} \to 0 \qquad \ldots \quad (3.7)$$

An application of Lemma 2 to (3.7) shows that

$$\sup_{\theta^*} \Sigma\, p_i\, [\pi_i(\theta^*) - p_i]^2 \to 0. \qquad \ldots \quad (3.8)$$

Since $\Sigma\, |p_i - \pi_i(\theta_0)| \to 0$, it follows easily from (3.8) that

$$\sup_{\theta^*} \Sigma\, \pi_i(\theta_0)[\pi_i(\theta^*) - \pi_i(\theta_0)]^2 \to 0. \qquad \ldots \quad (3.9)$$

We observe next, in consequence of A_2, that for any $\delta > 0$,

$$\inf_{|\theta - \theta_0| \geqslant \delta} \Sigma\, \pi_i(\theta_0)[\pi_i(\theta) - \pi_i(\theta_0)]^2 > 0. \qquad \ldots \quad (3.10)$$

It follows from (3.9) and (3.10) that with probability 1, for any $\delta > 0$, there exists an m (depending on c, δ, and the sequence $x_1, x_2 \ldots$) such that for all $n > m$ every approximate m.l. estimate θ^* is in the interval $(\theta_0 - \delta, \theta_0 + \delta)$. This establishes Theorem 1.

4. PROPERTIES OF M. L. EQUATION ESTIMATES

In this section, we state and prove some properties of the m.l. equation estimates and also establish the existence of m.l. estimates in the parametric case under certain additional assumptions.

Assumption A_3 : There exists a neighbourhood of θ_0 in which each $\pi_i(\theta)$ is differentiable with respect to θ.

Theorem 1. *Under assumptions* A_1, A_2 *and* A_3, *for all sufficiently large n with probability* 1,

(i) *m.l. estimates exist, and*

(ii) *every m.l. estimate is an m.l.e. estimate, and conversely.*

In other words, the class of m.l. estimates is non-empty and coincides with the class of m.l.e. estimates (definitions 2 and 4). It follows from Theorem 1 that this class is uniformly consistent.

To establish Theorem 2, choose and fix a $\delta > 0$ such that each $\pi_i(\theta)$ is differentiable (and therefore continuous) in $[\theta_0-\delta, \theta_0+\delta]$. Next, choose and fix a c, $0 < c < 1$, and consider the class of approximate m. l. estimates corresponding to c. It follows from Theorem 1 that with probability 1, for all sufficiently large n, every approximate m.l. estimate θ^* is in $(\theta_0-\delta, \theta_0+\delta)$. Consider a particular sequence $x_1, x_2, \ldots$ for which this is true, and a particular sufficiently large n. Let $L_n(\theta)$ denote the log likelihood at θ. If $\theta_1, \theta_2, \ldots$ is any sequence such that $L_n(\theta_k) \to \sup\limits_{\theta} \{L_n(\theta)\}$, we shall have $L_n(\theta_k) \geqslant \dfrac{\log c}{n} + \sup\limits_{\theta} \{L_n(\theta)\}$ for all sufficiently large k, since $0 < c < 1$, and hence $\theta_0-\delta < \theta_k < \theta_0+\delta$ for all sufficiently large k. It follows that the supremum of $L_n(\theta)$ equals the supremum with θ restricted to $[\theta_0-\delta, \theta_0+\delta]$, and is therefore attained in the latter interval. Thus m.l. estimates exist.

Now let $\hat{\theta}$ be any value of θ at which $L_n(\theta)$ is a maximum. Since then $\hat{\theta}$ is also an approximate m.l. estimate, we must have $\theta_0-\delta < \hat{\theta} < \theta_0+\delta$. Hence, as pointed out by Le Cam and Kraft (1956), the derivative of $L_n(\theta)$ must vanish at $\hat{\theta}$. Thus every m.l. estimate is a root of the likelihood equation, and therefore is an m.l.e. estimate. Conversely, every m.l.e. estimate must be an m.l. estimate, since the maximum of the likelihood with θ restricted to the roots of the likelihood equation has just been shown to equal the maximum likelihood with θ unrestricted. This completes the proof of Theorem 2.

The assumptions A_1, A_2 and A_3 under which the existence of consistent roots and their identification as m.l. estimates has been established here seem to be considerably weaker than those assumed by earlier authors. Also the argument does not depend on the number of unknown parameters so that the additional difficulty in establishing the existence of the roots of the m.l. equation present in some of the earlier proofs when the number of unknown parameters is more than one is avoided. However, proofs of other properties such as uniqueness of the m.l. estimate, and non-existence of two distinct consistent roots (Huzurbazar's theorem, 1948) seem to require additional assumptions. For example, if certain conditions on the second derivatives of $\pi_i(\theta)$ are satisfied, it is possible to deduce that, for some $r > 0$, with probability 1,

$$\max_{|\theta-\theta_0|<r} \{L_n''(\theta)\} < 0 \qquad \ldots \quad (4.1)$$

for all sufficiently large n, where dashes denote differentiation with respect to θ.

Let $\{t_{1n}\}$ and $\{t_{2n}\}$ be two consistent roots. If for some sequence and some n, $t_{1n} \neq t_{2n}$, it follows from $L_n'(t_{1n}) = L_n'(t_{2n}) = 0$ that there exists a

$$t_{3n} \in (t_{1n}, t_{2n}) \text{ with } L_n''(t_{3n}) = 0. \qquad \ldots \quad (4.2)$$

Since, with probability 1, the interval (t_{1n}, t_{2n}) is contained in $(\theta_0 - r, \theta_0 + r)$ for all sufficiently large n, it follows from (4.1) and (4.2) that we must have $t_{1n} = t_{2n}$ for all sufficiently large n. Thus the consistent root of the likelihood equation is unique in the sense that any two consistent roots coincide for all sufficiently large n. It follows in particular that the m.l.e. estimate is unique, and hence so is the m.l. estimate.

I wish to thank my colleague R. R. Bahadur for several useful discussions I had with him during the preparation of this paper.

REFERENCES

BAHADUR, R. R. (1958): Examples of inconsistency of maximum likelihood estimates. *Sankhyā*, **20**, 207.

HUZURBAZAR, V. S. (1948): The likelihood equation, consistency and the maximum of the likelihood function. *Ann. Eugenics*, **14**, 185.

KIEFER, J. AND WOLFOWITZ, J. (1956): Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Stat.*, **27**, 887.

KRAFT, C. AND LE CAM, L. (1956): A remark on the roots of the maximum likelihood equation. *Ann. Math. Stat.*, **27**, 1174.

KULLBACK, S. AND LEIBLER, R. A. (1951): On information and sufficiency. *Ann. Math. Stat.*, **22**, 79.

RAO, C. R. (1957): Maximum likelihood estimation for the multinomial distribution. *Sankhyā*, **18**, 139.

SCHEFFÉ, H. (1947): A useful convergence theorem for probability distributions. *Ann. Math. Stat.*, **18**, 434.

WALD, A. (1949): Note on the consistency of the maximum likelihood estimation. *Ann. Math. Stat.*, **20**, 595.

Paper received : June, 1958.

A REMARK ON STRONG MEASURABILITY

By V. S. VARADARAJAN

Indian Statistical Institute, Calcutta

SUMMARY. Assuming a proposition which is weaker than the continuum hypothesis, it is proved that every measurable map of a measure space into a metric space is strongly measurable, provided the σ–field in the basic space is generated by a denumerable sub-family.

Let $(\Omega, \mathcal{S}, \mu)$ be a probability space and let X be a metric space. Let $\mathcal{B}$ be the σ-field of the Borel subsets of X. Given a measurable map φ of Ω into X, it sometimes becomes important to know whether φ is strongly measurable (i.e.) whether φ is the almost everywhere limit of a sequence of simple mappings of Ω into X. This for instance is the case whenever X is a separable metric space or more generally, when the range of the map φ is a separable subset of X (Hille, 1948). The question was raised in a recent paper (Varadarajan, 1958) as to when measurable maps φ are strongly measurable. Here we give an interesting sufficient condition on the domain $(\Omega, \mathcal{S})$ which ensures that any measurable map of Ω into X has separable range (and hence is strongly measurable), X being an arbitrary metric space. The condition is that the cardinality of $\mathcal{S}$ be $\leqslant c$, the power of the continuum. In the proof of this result, we will assume the truth of proposition $\mathcal{P}$ below. This proposition is an immediate consequence of the continuum hypothesis, but does not seem to have been so far proved independently.

Proposition $\mathcal{P}$: *If* u *is an uncountable cardinal,* $2^u > c$.

We now state and prove the theorem.

Theorem : *Let* X *be an arbitrary metric space and* φ *a measurable map of* Ω *into* X. *If the cardinality of* $\mathcal{S}$ *is* $\leqslant c$, *and if proposition* $\mathcal{P}$ *is true, then* φ *has separable range.*

Proof : We observe first that there is no loss of generality in supposing that $\varphi[X] = X$, i.e. φ maps Ω onto X. Next, if A is any Borel subset of X, $\varphi^{-1}(A) \epsilon \mathcal{S}$ since φ is measurable. Finally, since $\varphi[X] = X$, $A \neq B$ implies $\varphi^{-1}(A) \neq \varphi^{-1}(B)$. Thus the correspondence : $A \rightarrow \varphi^{-1}(A)$ maps $\mathcal{B}$ in an one-one manner as a sub σ-field of $\mathcal{S}$, and hence the cardinality of $\mathcal{B}$ is $\leqslant c$. Now, it is known that this implies the separability of X, provided proposition $\mathcal{P}$ is true (Varadarajan and Ranga Rao, 1958). This completes the proof.

Remark : The condition on the cardinality of $\mathcal{S}$ is satisfied in many situations, e.g. when Ω is a Euclidean space or a separable metric space ($\mathcal{S}$ being then the class of Borel sets), or when $\mathcal{S}$ is generated by a denumerable sub-family.

REFERENCES

HILLE, E. (1948) : Functional analysis and semigroups. *Amer. Math. Soc.*, Colloquium publications.

VARADARAJAN, V. S. (1958) : On the convergence of sample probability distributions. *Sankhyā*, **19**, 23.

——— AND RANGA RAO, R. (1958) : On a theorem in metric spaces. *Ann. Math. Stat.* **29**, 612.

Paper received : May, 1958.

A USEFUL CONVERGENCE THEOREM

By V. S. VARADARAJAN

Indian Statistical Institute, Calcutta

SUMMARY. Let $X_1, X_2, \ldots$ be a sequence of k-dimensional random vectors. It is proved that in order that the distributions of the X_n should be convergent to a limit, it is necessary and sufficient that the distributions of $l(X_n)$ should converge to some limit for every linear function l.

The present paper is concerned with convergence of probability distributions on finite-dimensional Euclidean spaces. It is proved that for a given sequence $\{P_n\}$ of probability distributions to have a limit, it is necessary and sufficient that the one-dimensional distributions $\{P_n\lambda^{-1}\}$ have a limit for each linear function λ. This theorem seems useful in situations where the multidimensional theorem involves heavier computation. For example, this result ensures that in order that the sequence $\{P_n\}$ be asymptotically normal, it is necessary and sufficient that the sequence $\{P_n\lambda^{-1}\}$ be asymptotically normal for each linear function λ.

In what follows, X is a fixed k-dimensional space and S the σ-field of its Borel sets. Given a measure P on S and a measurable function λ on X, there exists a measure on the Borel sets of the real line, denoted by $P\lambda^{-1}$, such that $P\lambda^{-1}(E) = P\{\lambda^{-1}(E)\}$ for every Borel set E on the reals. If P_1 and P_2 are two measures on S, $P_1 = P_2$ when and only when $P_1\lambda^{-1} = P_2\lambda^{-1}$ for each linear function λ on X. A sequence $\{P_n\}$ of measure on S will be said to converge weakly to P if

$$\int_X g dP_n \to \int_X g dP$$

for every bounded continuous function g on X. This convergence is the classical convergence of P. Levy. It is possible to prove as in the one-dimensional case, that on the space of all measures on X this convergence arises through a metric (Varadarajan 1958, where a result of much greater generality is proved). The importance of this result lies in the fact that we can use simple topological arguments in many situations. We now state a lemma, which is true under more general situations (Kolmogorov and Prohorov 1954) but whose proof in the present case is very simple and well-known.

Lemma : *Let Γ be a set of probability measures on X. In order that Γ is conditionally compact* (i.e. $\bar{\Gamma}$ is compact) *it is necessary and sufficient that for each $\epsilon > 0$, there should exist a compact set K_ε such that*

$$P(K_\varepsilon) > 1-\epsilon$$

for all $P\epsilon\Gamma$.

We can now state and prove our theorem.

Theorem : *Let $\{P_n\}$ be a sequence of probability distributions on S. In order that there should exist a P such that $P_n \Longrightarrow P$, it is necessary and sufficient that for each linear function λ on X the sequence of one-dimensional distributions $\{P_n\lambda^{-1}\}$ converges to a limit.*

Proof : The necessity of the condition is obvious. We will prove the sufficiency now. Let $\Gamma = \{P_1, P_2, \ldots\}$. Let λ_i $(i = 1, 2, \ldots k)$ be the projection of X on its i-th component. Since $\{P_n\lambda_i^{-1}\}$ has a limit we can find numbers $a_{i\varepsilon}$ and $b_{i\varepsilon}$ such that

$$P_n\lambda_i^{-1}[a_{i\varepsilon}, b_{i\varepsilon}] > 1-\epsilon/k$$

for all n. If we now define $K_\varepsilon = \{(x_1, \ldots x_k) : a_{i\varepsilon} \leqslant x_i \leqslant b_{i\varepsilon} \text{ for } i = 1, 2, \ldots k\}$, we have

$$P_n(K_\varepsilon) > 1-\epsilon$$

for all n. This shows, in virtue of the lemma, that Γ is conditionally compact. If Γ has two limit points Q_1 and Q_2, then for any linear function λ, $Q_1\lambda^{-1}$ and $Q_2\lambda^{-1}$ are limit points of the sequence $\{P_n\lambda^{-1}\}$. This sequence is given to converge and hence it follows that $Q_1\lambda^{-1} = Q_2\lambda^{-1}$. Since λ is arbitrary, $Q_1 = Q_2$. Thus Γ can have only one limit point, say P. Since we have already shown that Γ is conditionally compact, it follows that $P_n \Longrightarrow P$. This completes the proof.

As an interesting application we have the following.

Corollary : *In order that $\{P_n\}$ may converge to a normal distribution on X, it is sufficient (and also necessary) that the sequence $\{P_n\lambda^{-1}\}$ converges to some normal distribution for each λ.*

References

Kolmogorov, A. N. and Prohorov, Yu. V. (1954) : *Bericht Uber Die Tagung Wahrscheinlichkeitsrechnung und Mathematische Statistik*, Berlin, 119.

Varadarajan, V. S. (1958) : Weak convergence of measures in separable metric spaces. *Sankhyā*, **19**, 15.

Paper received : May, 1958.

ON STATISTICS INDEPENDENT OF SUFFICIENT STATISTICS

By D. BASU

Indian Statistical Institute, Calcutta

SUMMARY. In an earlier paper (1955) the author stated that any statistic independent of a sufficient statistic must have the same distribution for all values of the unknown parameter. An example is given here to show that the proposition is not true in the generality stated above. Conditions under which the proposition is true are discussed.

1. Introduction

Let X be a random variable (sample) taking values in an arbitrary sample space $\mathscr{X}$ with the associated σ-field of measurable sets $\mathscr{A}$ and the family of probability measures $\{P_\theta\}$, $\theta \epsilon \Omega$. By a statistic $T = T(X)$ we mean a measurable characteristic of the sample X, i.e., T is an $\mathscr{A}$-$\mathscr{B}$ measurable transformation of the measurable space $(\mathscr{X}, \mathscr{A})$ into some measurable space $(\mathscr{T}, \mathscr{B})$. The family of induced (by the mapping T) probability measures on $\mathscr{T}$ is denoted by $\{P_\theta T^{-1}\}$, $\theta \epsilon \Omega$.

If $P_\theta T^{-1}$ is the same for all $\theta \epsilon \Omega$ then it is clear that an observation on the random variable T will be of no use for making any inference about the parameter θ. In this case we may say that the statistic T contains no information about the parameter θ. On the other hand if T be a sufficient statistic then we may say that T contains the whole of the information about θ that is contained in the sample X. Barring these two extreme situations it is not possible to make a general assessment of how much (or what proportion) of information is contained in a particular statistic. The author feels that the question 'How much information is contained in T ?' should be rephrased as 'How effective an observation on T is for making a particular inference about θ ?' Clearly the answer will depend on the kind of inference (tests of hypotheses, point or interval estimation etc.) that we wish to make and also on our idea (or criterion) of effectivity. An element of arbitrariness is bound to enter into any attempted definition of the amount of information in a statistic.

One interesting feature of Fisher's definition (1921) of the amount of information is that it is additive for independent statistics. That is, if T_1 and T_2 are any two statistics that are independent for every $\theta \epsilon \Omega$ then the amount of information in (T_1, T_2) is equal to the sum of the informations contained in T_1 and T_2 separately. This, however, does not appear to the author to be a necessary requirement for a satisfactory definition of information. It is possible to think of situations where T_1 and T_2 are equally informative (identically distributed for example) and are independent of one another but still, when an observation on T_1 is given, very little extra information will be supplied by an observation on T_2. For example, suppose we have a population whose distribution we know to be either $N(0, 1)$ or $N(5, 1)$. A single observation from the population will identify the true distribution with a great measure of certainty. Given one observation from the population very little extra information will be obtained from a second observation from the population. Surely the total information contained in two independent observations from the population is much less than twice that contained in a single observation. The following is a more extreme example.

Suppose it is known that a bag contains 10 identical balls numbered $\theta+1$, $\theta+2, \ldots, \theta+10$ where the unknown parameter θ takes any one of the values 0, 10, 20, 30, … . Suppose two balls are drawn one by one with replacement and let T_i be the number on the i-th ball drawn ($i = 1, 2$). Here T_1 and T_2 are identically distributed independent statistics and each is sufficient for θ. Given an observation on T_1 the distribution of T_2 gets completely specified and hence T_1 contains as much information as is contained in (T_1, T_2).

Independence of statistics is sometimes loosely interpreted as follows :— 'If the statistic T_2 is independent of T_1 then knowing what the realization of T_1 has been in a particular trial gives us no information about the possible realization of T_2 in the same trial.' When the probability measure on the sample space is only partially known the above interpretation of independence is no longer true. The example in the previous paragraph very forcefully brings this point out.

2. Statistics independent of a sufficient statistic

In the previous section we have given an example to show that a statistic can be independent of a sufficient statistic and still contain a great deal of information about the parameter. The example demonstrates that Theorem 1 in (1955) is not true in the generality stated there.[1] Under some mild restrictions, however, the theorem remains true.

---

[1] Mr. R. H. Farrell of the University of Illinois independently arrived at the conclusions contained in this section. In a letter to the author Mr. Farrell discussed the mistake in the proof of the theorem and gave an example very similar to the one considered above.

Let T be a sufficient statistic and let $A \epsilon \mathcal{A}$ be any fixed event that is independent of T. From the sufficiency of T it follows that there exists a $\mathcal{B}$-measurable real valued function $f(A|t)$ on $\mathcal{T}$(called the conditional probability of A given $T = t$) such that for any $B \epsilon \mathcal{B}$

$$P(A \cap T^{-1}B) = \int_B f(A|t) d\, P_\theta T^{-1} \text{ for all } \theta \epsilon \Omega.$$

From the independence of the event A and the statistic T it follows that for any $\theta \epsilon \Omega$,

$$f(A|t) = P_\theta(A) \qquad \dots \quad (2.1)$$

almost everywhere $[P_\theta T^{-1}]$ in t.

From (2.1) we cannot conclude that $P_\theta(A)$ is the same for all $\theta \epsilon \Omega$.

If the two measures $P_{\theta_1} T^{-1}$ and $P_{\theta_2} T^{-1}$ on $(\mathcal{T}, \mathcal{B})$ overlap, [i.e., for any set $B \epsilon \mathcal{B}$, $P_{\theta_1} T^{-1}(B) = 1$ implies that $P_{\theta_2} T^{-1}(B)$ is positive], then it is very easy to see that (2.1) implies the equality of $P_{\theta_1}(A)$ and $P_{\theta_2}(A)$.

Let us write $\theta_1 \Longleftrightarrow \theta_2$ if $P_{\theta_1} T^{-1}$ and $P_{\theta_2} T^{-1}$ overlap. The equality of $P_\theta(A)$ and $P_{\theta'}(A)$ can be deduced if there exists a finite number of parameter points $\theta_1, \theta_2, \dots \theta_k$ such that

$$\theta \Longleftrightarrow \theta_1 \Longleftrightarrow \theta_2 \dots \Longleftrightarrow \theta_k \Longleftrightarrow \theta'. \qquad \dots \quad (2.2)$$

We say θ and θ' are connected (by the statistic T) if there exists $\theta_1, \theta_2, \dots \theta_k$ satisfying (2.2).

Thus, we have the

Theorem : *If T be a sufficient statistic and if every pair of θ's in Ω are connected (by T), then any event A independent of T has the same probability for all* $\theta \epsilon \Omega$.

As a corollary we at once have that under the conditions of the above theorem any statistic T_1 independent of the sufficient statistic T contains no information about the parameter.

REFERENCES

BASU, D. (1955) : On statistics independent of a complete sufficient statistic. *Sankhyā*, **15**, 377.

FISHER, R. A. (1921) : The mathematical foundations of theoretical statistics. *Phil. Trans. Royal Soc.* A, **222**, 309.

Paper received : May, 1958.

BUREAU OF EDNL. &
LIBRARY

ON CONFIGURATIONS AND NON-ISOMORPHISM OF SOME INCOMPLETE BLOCK DESIGNS

By M. ATIQULLAH

University of Dacca, Pakistan

SUMMARY. A necessary and sufficient condition for the existence of a particular class of Balanced Incomplete Block (BIB) designs has been obtained in this paper. It is established that a BIB design with parameters:

$$V = n(n-1)/2;\quad b = n(n+1)/2;\quad k = (n-1);\quad r = (n+1),\quad \lambda = 2,$$

can be arranged into two groups D_1 and D_2, where D_1 and D_2 are two associate Partially Balanced Incomplete Block (PBIB) designs with parameters depending on n. The converse is also proved. The existence of the design D is thus made to depend on the existence of the design D_2, the impossibility of which for certain values of n immediately establishes the impossibility of the corresponding design D.

Besides, an attempt has been made to develop a method for construction as well as for determination of the non-isomorphic solutions for certain PBIB designs. Some new solutions are obtained and in some cases several designs are demonstrated to be impossible combinatorially.

1. INTRODUCTION

The principle of Incomplete Block designs was evolved by Yates (1936) on which extensive work has been carried on in recent years. The problem of enumeration of all possible non-isomorphic solutions for Balanced Incomplete Block (BIB) designs were discussed by Fisher (1940), Cox (1940), Husain (1945) and (1945a). Bose and Nair (1939) introduced a wider class of designs, viz., Partially Balanced Incomplete Block (PBIB) design which includes as a special case the BIB designs and the square lattices. A generalisation by Nair and Rao (1942), resulted in the inclusion of cubic and other higher dimensional lattices as special cases. Recently a distinct advance has been made by Bose and his co-workers in U.S.A. in establishing more important properties of PBIB designs and in some cases discovering combinatorial solutions for a number of designs. But the problem of testing the isomorphism of the solutions of a PBIB design remains to be attacked, as also the problem of investigating the totality of non-isomorphic solutions.

In the present paper, a necessary and sufficient condition for the existence of a particular class of BIB design has been arrived at. Besides, an attempt has been made to develop a method for construction as well as for determination of the non-isomorphic solutions for certain PBIB designs. This method is based on the concept of "chains" which was originally introduced by Husain (1945) but it was exclusively devoted to the case of a special type of symmetrical BIB design only. The ideas have been modified by the author and its application extended to certain PBIB designs as well as to a particular type of BIB designs. Also, some new solutions are obtained and in some cases several designs are demonstrated to be impossible combinatorially.

2. RESOLVABILITY AND IMPOSSIBILITY OF SOME INCOMPLETE BLOCK DESIGNS

Definition : An Incomplete Block design with v varieties each replicated r times in b blocks of k experimental units each ($k < v$) and no variety placed more than once in the same block is said to be a PBIB design if the following conditions hold:

(i) There can be established an association relationship between any two varieties such that (a) two varieties are either 1st, 2nd... or m-th associates, (b) each variety has exactly n_i, i-th associates—($i = 1, 2, \ldots, m$) and (c) given any two varieties which are i-th associates the number of varieties common to the j-th associates of the first, and k-th associates of the second is p^i_{jk} and this number is independent of the pair of varieties with which we start. Also, $p^i_{jk} = p^i_{kj}$, $(i, j, k = 1, 2, \ldots, m)$.

(ii) Two varieties which are i-th associates occur together in exactly λ_i blocks. Without loss of generality it is generally assumed that $\lambda_i \geqslant \lambda_{i+1}$ which implies that the λ's need not all be distinct. When all the λ's are equal, we get a BIB design. In the so-called symmetrical Incomplete Block design, $v = b$ (or equivalently $r = k$). It is known that the following conditions are satisfied by the parameters of $v, b, r, k, \lambda_1, \lambda_2, \ldots, \lambda_m, n_1, n_2, \ldots, n_m, p^i_{jk}$ $(i, j, k = 1, 2, \ldots, m)$ of the design.

$$vr = bk,$$

$$v-1 = \sum_{i=1}^{m} n_i,\ r(k-1) = \Sigma\, n_i \lambda_i,$$

$$\sum_{k=1}^{m} p^i_{jk} = \begin{cases} n_i - 1 & \text{if } i = j, \\ n_j & \text{if } i \neq j, \end{cases}$$

$$n_i\, p^i_{jk} = n_j\, p^j_{ik} = n_k\, p^k_{ji} \quad (i, j, k = 1, 2, \ldots, m). \qquad \ldots \quad (2.1)$$

If $m = 2$, then clearly

$$vr = bk,$$

$$v-1 = n_1 + n_2,\ r(k-1) = n_1\lambda_1 + n_2\lambda_2,$$

$$p^1_{11} + p^1_{12} + 1 = p^2_{11} + p^2_{12} = n_1,$$

$$p^1_{21} + p^1_{22} = p^2_{21} + p^2_{22} + 1 = n_2,$$

$$n_1 p^1_{12} = n_2 p^2_{11} \text{ and } n_1 p^1_{22} = n_2 p^2_{12}. \qquad \ldots \quad (2.2)$$

Furthermore, if values are assigned to the parameters of the first kind ($v, b, r, k, \lambda_1, \lambda_2, n_1$ and n_2) satisfying (2.2), then there is one independent parameter of the second kind (p^i_{jk} for $i, j, k = 1, 2$). The parameters of the second kind is customarily exhibited as elements of two symmetric matrices

$$p^1_{jk} = \begin{pmatrix} p^1_{11} & p^1_{12} \\ p^1_{21} & p^1_{22} \end{pmatrix},\ p^2_{jk} = \begin{pmatrix} p^2_{11} & p^2_{12} \\ p^2_{21} & p^2_{22} \end{pmatrix}.$$

Bose and Shimamoto (1952) gave among different types of PBIB designs for two associate classes a special type of design TSLB (Triangular Singly Linked Block) which has the following parameters;

$$v = \frac{n(n-1)}{2},\ b = n, r = 2, k = n-1, \lambda_1 = 1, \lambda_2 = 0,$$

$$n_1 = 2n-4, \quad n_2 = (n-2)(n-3)/2,$$

$$p^1_{jk} = \begin{pmatrix} n-2 & n-3 \\ n-3 & (n-3)(n-4)/2 \end{pmatrix}$$

$$p^2_{jk} = \begin{pmatrix} 4 & 2n-8 \\ 2n-8 & (n-4)(n-5)/2 \end{pmatrix}. \qquad \ldots \quad (2.3)$$

It will be shown that this design forms a part of a BIB design with the same k having parameters,

$$v = \frac{n(n-1)}{2}, \quad b = \frac{n(n+1)}{2}, \quad k = n-1, \quad r = n+1, \quad \lambda = 2. \qquad \ldots \quad (2.4)$$

For future discussion; any block chosen arbitrarily of a design under study may be called an initial block; it may be denoted by the letter U and its varieties by different θ's and the non-initial varieties (those not contained within U) by different ψ's. Small letter u_t will stand for the number of blocks having $t\,(= 0, 1, 2, \ldots, k)$ varieties common with U and the capital letter U_t for the set of blocks each having $t\,(=0, 1, 2, \ldots, k)$ varieties common with U. The BIB design (2.4) and the TSLB design (2.3) will be shortly designated by W and Γ respectively.. It is well-known that a design Γ always exists.

Theorem 1 : *A necessary and sufficient condition for the existence of a BIB design W for $n \geqslant 4$ is that there exists a symmetrical PBIB design with the following parameters :*

$$v = b = \frac{n(n-1)}{2}, \quad r = k = n-1, \quad \lambda_1 = 2, \lambda_2 = 1.$$

$$n_1 = (n-2)(n-3)/2, \quad n_2 = 2n-4$$

$$p^1_{jk} = \begin{pmatrix} (n-4)(n-5)/2 & 2n-8 \\ 2n-8 & 4 \end{pmatrix}$$

$$p^2_{jk} = \begin{pmatrix} (n-3)(n-4)/2 & n-3 \\ n-3 & n-2 \end{pmatrix}. \qquad \ldots \quad (2.5)$$

Proof : Let π denote a symmetrical PBIB design (2.5) and suppose this exists. Then a BIB design W can be obtained by the juxtaposition of π and Γ. For, a variety which occurs $n-1$ times in π must necessarily occur twice in Γ. Also, a pair of varieties which occurs once or twice in π must occur once or zero times respectively, in Γ. It implies that the same varieties which are first associates of a particular variety in π become second associate in Γ; and so the value of p^1_{11} in π corresponds to p^2_{22} in Γ. Similarly other parametric values involved in Γ can be determined by suitable shuffling as shown in (2.3) of the parametric values of π. This proves the necessary condition.

For sufficiency, let U be a block of an existing BIB design W, and let there be u_t blocks U_t, $(t = 0, 1, \ldots, n-1)$. Husain (1945) showed that $u_1 = 2(n-1)$ and $u_2 = (n-1)(n-2)/2$, and other U_t's are zero. He listed certain general properties of the set of U_1 blocks which have not as yet been thoroughly tried to establish their resolvability into two unique sets of singly linked blocks. This however follows from the results obtained by Hall and Connor (1953). We get the arrangement of $u_1 = 2(n-1)$, blocks into the unique sets, S_1 and S_2, each of $n-1$ blocks and having the properties that (i) a block of S_1 or S_2 has got one distinct θ-variety common with U; (ii) every ψ-variety occurs just twice in S_1 or S_2; (iii) any two blocks of S_1 or S_2 have exactly one ψ-variety in common ; (iv) each block of S_1 has one variety common with a distinct block of S_2 and two varieties common with $n-2$ other blocks of S_2 and vice-versa. The properties (i), (ii) and (iii) establish the fact that the block U together with any one set, say S_1, gives a design Γ, and therefore the set of blocks U_2 together with the other set S_2 must constitute a design π.

Corollary 1 : *Two symmetrical PBIB designs π can be obtained from a known BIB design W.*

It is evident from the above results that the set S_1 together with the set of blocks U_2 and also the set S_2 together with the same set U_2 yield two π designs.

Corollary 2 : *A necessary and sufficient condition for the existence of a symmetrical BIB design with the parameters, $r = n+1$ and $\lambda = 2$ is that there exists a symmetrical PBIB design π.*

As shown above, a BIB design W can be obtained from a known design π by the juxtaposition of π and Γ. This according to Hall and Connor (1953) can always be fringed to the above symmetrical BIB design. Conversely, let S be the set of blocks containing any particular variety in an existing symmetrical BIB design with $\lambda = 2$, and let $\bar{S}$ be the set of the remaining blocks. We select any block from S and omit all the varieties contained within it from the other blocks of S and $\bar{S}$. It is found that the residual configuration obtained from S and $\bar{S}$ are Γ and π respectively.

Theorem 2 : *A necessary condition for the existence of a symmetrical PBIB design, π, for $n \geqslant 4$, where $\frac{n(n-3)}{2}$ is odd, is that r must be a perfect square.*

Proof : Let the incidence matrix of the design π be $N = (n_{ij})$, where $n_{ij} = 1$ or 0 according as the i-th variety ($i = 1, 2, \ldots, v$) does or does not occur in the j-th block ($j = 1, 2, \ldots, v$). Then, the value of the determinant NN' follows from the results due to Connor and Clatworthy (1954),

$$|NN'| = r^2(r-z_1)^{\alpha_1}(r-z_2)^{\alpha_2}; \quad \alpha_1+\alpha_2 = v-1 \qquad \ldots \quad (2.6)$$

where z_1 and z_2 are distinct and α_1 and α_2 are positive integers. These constants are given by,

$$Z_1 = \tfrac{1}{2}[(\lambda_1-\lambda_2)(-\gamma-\sqrt{\Delta})+(\lambda_1+\lambda_2)],$$

$$Z_2 = \tfrac{1}{2}[(\lambda_1-\lambda_2)(-\gamma+\sqrt{\Delta})+(\lambda_1+\lambda_2)],$$

$$\alpha_1 = [(v-1)(-\gamma+\sqrt{\Delta}+1)-2n_1]/2\sqrt{\Delta},$$

$$\alpha_2 = [(v-1)(\gamma+\sqrt{\Delta}+1)-2n_2]/2\sqrt{\Delta}, \qquad \ldots \quad (2.7)$$

where $\gamma = P^2_{12}-P^1_{12}\ ;\ \beta = P^2_{12}+P^1_{12}$

and $\Delta = \gamma^2+2\beta+1.$

After simplification, we get $z_1 = 0, \quad z_2 = n-2;$

$$\alpha_1 = n(n-3)/2; \qquad \alpha_2 = n-1;$$

so that we have,

$$|NN'| = r^{(n^2-3n+4)/2}. \qquad \ldots \quad (2.8)$$

Since $|NN'|$ is an integral square, it follows, if $\dfrac{n^2-3n}{2}$ is odd, r must be a perfect square.

Corollary 2 : *The symmetrical PBIB design π corresponding to $n = 6, 9, 13, 14,$ and 18 are impossible combinatorially which lead eventually to the non-existence of the following BIB designs as well as of their corresponding symmetrical BIB designs :*

(i) $v = 15, \quad b = 21, \quad r = 7, \quad k = 5, \quad \lambda = 2$
(ii) $v = 36, \quad b = 45, \quad r = 10, \quad k = 8, \quad \lambda = 2$
(iii) $v = 78, \quad b = 91, \quad r = 14, \quad k = 12, \quad \lambda = 2$
(iv) $v = 91, \quad b = 105, \quad r = 15, \quad k = 13, \quad \lambda = 2$
(v) $v = 153, \quad b = 171, \quad r = 19, \quad k = 17, \quad \lambda = 2.$

The impossibility of (i) was first demonstrated by Nandi (1946) by a process of complete enumeration and the impossibility of (ii) was proved by Connor (1952) with the help of a structural matrix. The impossibility of the symmetrical BIB designs corresponding to them was established by Shrikhande (1950). The above is an alternative simple proof of these designs.

Theorem 3 : *The dual design (i.e. the design obtained by taking the blocks as varieties and the varieties as blocks) of a symmetrical PBIB design, π, is also a symmetrical PBIB design satisfying the same set of parameters* (2.5).

Proof : Let D be any block of π. Since any two blocks of π have one variety or two varieties in common, let there be d_1 blocks D_1 and d_2 blocks D_2 in π.

Then

$$d_1+d_2 = \frac{(n-2)(n+1)}{2}$$

$$d_1+2d_2 = (n-1)(n-2).$$

Hence $$d_1 = 2(n-2) \text{ and } d_2 = (n-2)(n-3)/2.$$

In the dual design, therefore, with respect to any given variety, the other varieties can be divided into two groups of n_1 and n_2 items, respectively, where $n_1 = (n-2)(n-3)/2$ and $n_2 = 2n-4$, such that the varieties of the first group occur $\lambda_1(= 2)$ times and those of the second group occur $\lambda_2(= 1)$ times with the given variety.

Let E be a member of D_1 and let there be e_1 blocks E_1 and e_2 blocks E_2 in set D_1. Also let there be f_1 blocks E_1 and f_2 blocks E_2 in the set D_2. Then it is found that the number of blocks having one variety common with D and simultaneously with E is $n-2$, i.e., $e_1 = n-2$. Again, the number of blocks having one variety common with D and simultaneously having two varieties common with E is $n-3$, i.e., $e_2 = n-3$. Also, $f_1 = n-3$ and $f_2 = (n-3)(n-4)/2$.

In the dual design we shall have therefore, $P^2_{22} = n-2$; $P^2_{12} = n-3$ and $P^2_{11} = (n-3)(n-4)/2$. It can be verified that the other parameters also are satisfied by the dual design.

3. TEST FOR ISOMORPHISM AND THE TOTALITY OF NON-ISOMORPHIC SOLUTIONS OF CERTAIN PARTIALLY BALANCED DESIGNS

Two solutions satisfying the same set of parameters are called isomorphic if one can be derived from the other by renaming the varieties (the order of the blocks or of the varieties in a block being immaterial). To find suitable transposition or permutation of varieties which can convert one solution into another is rather tedious and often difficult, in case the number of blocks or the varieties are large and so it is desirable to find some methods based on certain criteria which can facilitate isomorphism of solutions. A method of "Chain" has been introduced to study the isomorphism for the solutions of PBIB designs (2.5) and (4.1) and also for BIB design (2.4). This method can be profitably used in connection with a wider class of designs with necessary modifications, and modified rules for 'Consistency' of 'Chains'. The actual meaning of "Chain" in a particular case will be obvious from the text. However, the underlying ideas are explained here in brief with reference to designs having two associate classes for λ's not exceeding 2.

We try to select a group of varieties having the property that the different blocks of a design may be specified uniquely by the varieties contained within them. We may call such varieties the X-varieties and the remaining varieties the Y-varieties. The X-varieties may be elements of any one particular block; but in general they may not be so. Once the blocks have been specified by the X-varieties, the location of a Y-variety may thereafter be described by naming the X-varieties descriptive of the blocks containing the particular Y-variety.

If a particular Y-variety, say Y_0, occurs in m blocks, each containing one different $X_j (j = 1, 2, \ldots, m)$, the location of Y_0 may be written shortly as $Y_0 : \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$ and the symbolic representation $\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_m \end{bmatrix}$ may be called a "Link". Each element of a "Link" identifies one distinct block.

If Y_0 occurs in q blocks each containing one different pair of X-varieties, say $X_1 X_2$; X_2X_3; X_3X_4, for $q = 4$, the location of Y_0 in such blocks may be written shortly as $Y_0 : (X_1X_2X_3X_4)$ and the symbolic representation $(X_1X_2X_3X_4)$ may be called a "Ring". A 'Ring' contains at least two elements. Each pair of consecutive elements of a ring identifies one distinct block. A ring can begin from its first element and end at its last element and it can be read either forward or backward. In case it has more than three elements, its first and its last elements may be identical but the intermediate elements must be different. For example, a ring of the type $(X_1X_2X_3X_1)$ or $(X_1X_2X_3X_4)$ can be read in the first case as X_1X_2; X_2X_3; X_3X_1 or X_1X_3; X_3X_2; X_2X_1 and in the second case as X_1X_2; X_2X_3; X_3X_4 or X_4X_3; X_3X_2; X_2X_1 and thus it represents in either case three blocks. A ring having two elements only may be called a 'simple ring'.

Thus the location of Y_0 in a design with two associate classes for which λ's do not exceed the numerical value 2, may be described by a "Chain" consisting of

(i) one or more "Links"; or

(ii) one or more "Rings"; or

(iii) both "Links" and "Rings".

Let Y_0 have N_1 first associates and N_2 second associates among the other Y-varieties. The value of N_1 and N_2 are determined by the parameters of a design under study.

Then a set of 'chains' may be called consistent, provided

(i) between any two chains there be exactly λ_1 or λ_2 blocks common,

(ii) any particular 'chain' shall have exactly λ_1 blocks common with N_1 chains and just λ_2 blocks common with N_2 remaining chains.

Since the pattern of a 'chain' cannot be altered by any permutation of the X-varieties, two solutions in order to be isomorphic must necessarily have similar pattern of chains, each type occurring with the same frequency. When this necessary condition is satisfied, we can try to identify one type of chain of a solution with the corresponding type of chain of the other by a suitable permutation. If this succeeds and the same permutation works for all other types, then the Y-varieties corresponding to the identified chains can be identified.

Our main object is to study whether the two solutions of π, as discussed in earlier chapter obtained from the same existing BIB design W are isomorphic. Firstly, the uniqueness of a Γ design will be proved. If so, all possible non-isomorphic solutions of a single BIB design W can be detected by writing out these designs in a standard form such that each contains an identical Γ design; and then testing for isomorphism the π designs left out. It will be shown that a known design W yields two non-isomporhic π designs.

Uniqueness of a Γ *design* 1. It is already known that a Γ design always exists. For, it can be written down by taking an $n\times n$ square and filling the $\frac{n(n-1)}{2}$ positions above the leading diagonal by the different varieties in such a way that the positions in the leading diagonal are left blank, while the positions below this are filled so that the scheme is symmetrical with respect to the diagonal. This design can be also constructed by following the principle of 'chains'. This method facilitates, to a great extent, in establishing the uniqueness of Γ.

Let any block of Γ be represented by U and the set of other remaining blocks by $\bar{\Gamma}$. Since a variety occurs just twice in the n blocks and a pair occurs at most in one block, every block of $\bar{\Gamma}$ can be identified by one distinct θ-variety of U and a ψ-variety can be identified by the two θ-varieties occurring in the two blocks containing it. Clearly, the group of θ-varieties satisfy the property of X-varieties, and so the two θ-varieties, say θ_1 and θ_2 defining a particular ψ-variety say ψ_0, may be written as $\psi_0 : \begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}$ and, according to our definition, $\begin{bmatrix}\theta_1\\ \theta_2\end{bmatrix}$ may be called a "Link". Again let the ψ_0 have N_1 first associates and N_2 second associates among the other ψ-varieties. As each ψ-variety occurs twice in $\bar{\Gamma}$ we have by the parametric relations (2.1),

$$N_1+N_2 = n(n-3)/2,$$

$$N_1\,\lambda_1+N_2\lambda_2 = 2(n-3),$$

whence, $$N_1 = 2(n-3) \text{ and } N_2 = (n-3)(n-4)/2.$$

Now, a set of links may be called consistent if any link has got exactly one θ-element common with $2(n-3)$ links and no element common with the remaining $(n-3)(n-4)/2$ links. There are as many different links as there are ψ-varieties; once all the links are written down, it is found that the rule of consistency is automatically satisfied; and thus, when all the ψ-varieties are assigned to the links a solution is reached; and conversely if a Γ design be given then with each of the ψ-varieties is associated one distinct link.

Because of one to one correspondence between the links and the ψ-varieties the total number of the Γ designs that can be written out from a given initial block is $\left(\frac{(n-1)(n-2)}{2}\right)$! the number of ways in which the ψ-varieties can be assigned to the links or the links to the ψ-varieties. Evidently all these $\left(\frac{(n-1)(n-2)}{2}\right)$! solutions are isomorphic.

π "*chains*" 2. As proved in Theorem 1, a design W for $n \geqslant 4$, if it exists, is resolvable uniquely into two configurations, Γ and π where π, with reference to a block U of Γ, is constituted by the set of blocks S_2 and U_2, and the properties of S_2 are essentially those of $\bar{\Gamma}$ blocks, it follows that with each of the ψ-varieties of S_2 can be associated one distinct link. Since every block of the set U_2 contains one different pair of θ-varieties, the location of a ψ-variety in this set can be expressed by naming the corresponding θ-pairs. We can designate such pairs concisely by means of one or more "rings". In short, the location of a ψ in π can be described by a "chain" consisting of one 'link' and one or more "Rings". Such a chain may be called a π-chain.

Let any particular ψ-variety of π have N_1 first associates and N_2 second associates among the other ψ-varieties. Remembering that a ψ occurs twice in S_2 and $n-3$ times in U_2, we have by the parametric relation (2.1),

$$N_1+N_2 = n(n-3)/2$$

$$N_1\lambda_1+N_2\lambda_2 = 2(N-3)+(n-3)(n-4)$$

whence $$N_1 = \frac{(n-3)(n-4)}{2} \text{ and } N_2 = 2n-6.$$

Thus for $\lambda_1 = 2$ and $\lambda_2 = 1$, a set of π-chains may be called consistent, if any chain has got exactly two blocks common with $\frac{(n-3)(n-4)}{2}$ chains and just one block common with the remaining $2n-6$ chains. The number of blocks common between the chains are counted by noting down the common elements between the links and the consecutive pair of elements common between the rings of the two chains. It is thus clear that if we can find a set of consistent π-chains we can get a design π and conversely, if there be given a design π there will correspond to each ψ-variety one distinct chain of a consistent set.

β-chains 3. Given a design Γ, a design W can be constructed from it if the blocks of the set S_2 and U_2 can be determined. It follows from the above arguments that the location of a ψ-variety in a BIB design W can be described by a "chain" consisting of two links (one each for Γ and S_2) and one or more rings (for U_2). We may call such chains the "β-chains". The fact that every ψ occurs twice with the θ's for $\lambda = 2$, ensures that each of the θ's must occur twice in a β-chain, remembering that the intermediate elements of a ring are counted twice. Two β-chains would be consistent if there are exactly two blocks common between them. The number of blocks common between two β-chains are counted by noting down the common elements between the links corresponding to Γ, the common elements between the links corresponding to S_2, and the consecutive pair of elements common between the rings. Thus a BIB design W can be constructed, if it is possible to find a set of β-chains such that every pair of them is consistent, and conversely if a solution be given, then, with each of the ψ-varieties is associated one distinct chain of a set, every pair of which is consistent.

For a given Γ, we determine by trial a set of $n-1$ blocks S_2 such that the blocks of S_2 and Γ satisfy the four properties, as listed in Theorem 1. Once the set S_2 is found, the construction of consistent β-chains becomes easy. As an illustration, consider a design Γ for $n = 4$. Let the different θ-varieties be represented by the integers 1, 2, 3, and the ψ-varieties by the numbers 4, 5, 6. The links corresponding to the ψ-varieties, 4, 5, 6 of Γ are given by : $\begin{bmatrix}1\\2\end{bmatrix}$; $\begin{bmatrix}1\\3\end{bmatrix}$; $\begin{bmatrix}2\\3\end{bmatrix}$. We get by trial a set of blocks S_2 whose ψ-varieties, 4,5, and 6, have got the links, $\begin{bmatrix}2\\3\end{bmatrix}$, $\begin{bmatrix}1\\2\end{bmatrix}$, $\begin{bmatrix}1\\3\end{bmatrix}$ respectively. The location of the ψ in Γ and S_2 taken together, is conveniently represented as :

$$4 : \begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix};\ 5 : \begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix};\ 6 : \begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}.$$

To get β-chains the varieties 4, 5, 6 must occur in the blocks (13), (23), and (12) respectively. Thus a set of consistent β-chains are given by:

$$4 : \begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}(13);\ 5 : \begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}(23);\ 6 : \begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}(12).$$

Any existing π design can be expressed in terms of its π-chains and to construct from it a design W, it is only necessary to determine one other set of links which completes the β-chains. This is accomplished, keeping in view that each element in the β-chain must occur just twice. It is to be noted that the whole set of β-chains thus obtained from a set of consistent π-chains forms a consistent set.

The design π for $n = 4$ has a unique solution. It is well-known that the design W for $n = 4$ admits of one independent solution. Consequently, two π-designs obtained from the above solution of (3) are given as follows :

$$\text{(a)}\quad 4 : \begin{bmatrix}1\\2\end{bmatrix}(13);\quad 5 : \begin{bmatrix}1\\3\end{bmatrix}(23);\quad 6 : \begin{bmatrix}2\\3\end{bmatrix}(12)$$

$$\text{(b)}\quad 4 : \begin{bmatrix}2\\3\end{bmatrix}(13);\quad 5 : \begin{bmatrix}1\\2\end{bmatrix}(23);\quad 6 : \begin{bmatrix}1\\3\end{bmatrix}(12).$$

We omit the second set of links from the β-chains of the solution in (3) and get the above design (a). Similarly, by omitting the first set of links the design (b) is obtained. It is obvious that the two designs, (a) and (b), are isomorphic. For, an interchange of the θ-varieties, 2 and 3, brings the two sets of chains to identify themselves, and as a result ψ-varieties, 4 and 6, get interchanged. This establishes the uniqueness of π for $n = 4$.

The design π for $n = 5$ have five independent solutions. Consider a design Γ for $n = 5$, in which an initial block contains the θ-varieties (1, 2, 3, 4) and in the standard setting let the links for the ψ-varieties, 5, 6, 7, 8, 9, and 10 be respectively given by:

$$\begin{bmatrix}1\\2\end{bmatrix};\ \begin{bmatrix}1\\3\end{bmatrix};\ \begin{bmatrix}1\\4\end{bmatrix};\ \begin{bmatrix}2\\3\end{bmatrix};\ \begin{bmatrix}2\\4\end{bmatrix};\ \begin{bmatrix}3\\4\end{bmatrix}.$$

It has been found that the totality of the designs W for $n = 5$ obtained with the help of β-chains fall, broadly speaking, in three types of classes, C_1, C_2 and C_3 as shown below:

TABLE 1. NON-ISOMORPHIC SET OF β-CHAINS

non-initial varieties	β-chains		
	C_1	C_2	C_3
5:	$\begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$ (12)(34)	$\begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$ (134)	$\begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$ (134)
6:	$\begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$ (13)(24)	$\begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$ (124)	$\begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}$ (142)
7:	$\begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}$ (14)(23)	$\begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}$ (14)(23)	$\begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$ (123)
8:	$\begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$ (23)(14)	$\begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$ (14)(23)	$\begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$ (14)(23)
9:	$\begin{bmatrix}2\\4\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}$ (24)(13)	$\begin{bmatrix}2\\4\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}$ (13)(24)	$\begin{bmatrix}2\\4\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}$ (13)(24)
10:	$\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$ (12)(34)	$\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$ (12)(34)	$\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$ (12)(34)

A design of C_1-type has got all of its six chains associated with double simple rings. A design of C_2-type possesses four chains with double simple ring and of C_3-type has three chains with double simple rings. Nandi (1945) showed that the BIB design W for $n = 5$ admits of three independent solutions classified by him as $[\alpha_{11}]$, $[\alpha_3]$ and $[\beta_{21}]$. These can be identified with our C_2, C_3 and C_1 respectively. A special feature of a C_1-type design is that it provides a unique π design. But a C_2-type and C_3-type each provides separately for two indepedent π designs.

Hence we get in all five independent solutions as listed below:

TABLE 2. NON-ISOMORPHIC SET OF π-CHAINS

non-initial varieties	π_0	π_1	π_2	π_3	π_4
5:	$\genfrac{}{}{0pt}{}{1}{2}$ (12)(34)	$\genfrac{}{}{0pt}{}{1}{2}$ (134)	$\genfrac{}{}{0pt}{}{2}{3}$ (134)	$\genfrac{}{}{0pt}{}{1}{2}$ (134)	$\genfrac{}{}{0pt}{}{2}{4}$ (134)
6:	$\genfrac{}{}{0pt}{}{1}{3}$ (13)(24)	$\genfrac{}{}{0pt}{}{1}{3}$ (142)	$\genfrac{}{}{0pt}{}{2}{3}$ (142)	$\genfrac{}{}{0pt}{}{1}{3}$ (124)	$\genfrac{}{}{0pt}{}{3}{4}$ (124)
7:	$\genfrac{}{}{0pt}{}{1}{4}$ (14)(23)	$\genfrac{}{}{0pt}{}{1}{4}$ (123)	$\genfrac{}{}{0pt}{}{3}{4}$ (123)	$\genfrac{}{}{0pt}{}{1}{4}$ (14)(23)	$\genfrac{}{}{0pt}{}{2}{3}$ (14)(23)
8:	$\genfrac{}{}{0pt}{}{2}{3}$ (14)(23)	$\genfrac{}{}{0pt}{}{2}{3}$ (14)(23)	$\genfrac{}{}{0pt}{}{1}{4}$ (14)(23)	$\genfrac{}{}{0pt}{}{2}{3}$ (14)(23)	$\genfrac{}{}{0pt}{}{1}{4}$ (14)(23)
9:	$\genfrac{}{}{0pt}{}{2}{4}$ (13)(24)	$\genfrac{}{}{0pt}{}{2}{4}$ (13)(24)	$\genfrac{}{}{0pt}{}{1}{3}$ (13)(24)	$\genfrac{}{}{0pt}{}{2}{4}$ (13)(24)	$\genfrac{}{}{0pt}{}{1}{3}$ (13)(24)
10:	$\genfrac{}{}{0pt}{}{3}{4}$ (12)(34)	$\genfrac{}{}{0pt}{}{3}{4}$ (12)(34)	$\genfrac{}{}{0pt}{}{1}{2}$ (12)(34)	$\genfrac{}{}{0pt}{}{3}{4}$ (12)(34)	$\genfrac{}{}{0pt}{}{1}{2}$ (12)(34)

The design π_2 or π_1 is obtained by omitting the first or the second set of links from the β-chains of C_2. Similarly π_4 or π_3 is obtained from C_3. It is found that no substitution brings π_1 to π_2 or π_3 to π_4. Moreover, the design π_0 is self-dual and so is π_3. The others are isomorphic with their respective duals. Hence we have in all five non-isomorphic solutions.

π design for $n = 8$. Consider a design Γ, with its initial block containing the θ-varieties, 1, 2, 3, 4, 5, 6, 7. Let the ψ-varieties associate with their links be given as follows:

$$8 : \begin{bmatrix}1\\2\end{bmatrix};\ 9 : \begin{bmatrix}1\\3\end{bmatrix};\ 10 : \begin{bmatrix}1\\4\end{bmatrix};\ 11 : \begin{bmatrix}1\\5\end{bmatrix};\ 12 : \begin{bmatrix}1\\6\end{bmatrix};\ 13 : \begin{bmatrix}1\\7\end{bmatrix};\ 14 : \begin{bmatrix}2\\3\end{bmatrix};$$

$$15 : \begin{bmatrix}2\\4\end{bmatrix};\ 16 : \begin{bmatrix}2\\5\end{bmatrix};\ 17 : \begin{bmatrix}2\\6\end{bmatrix};\ 18 : \begin{bmatrix}2\\7\end{bmatrix};\ 19 : \begin{bmatrix}3\\4\end{bmatrix};\ 20 : \begin{bmatrix}3\\5\end{bmatrix};\ 21 : \begin{bmatrix}3\\6\end{bmatrix};$$

$$22 : \begin{bmatrix}3\\7\end{bmatrix};\ 23 : \begin{bmatrix}4\\5\end{bmatrix};\ 24 : \begin{bmatrix}4\\6\end{bmatrix};\ 25 : \begin{bmatrix}4\\7\end{bmatrix};\ 26 : \begin{bmatrix}5\\6\end{bmatrix};\ 27 : \begin{bmatrix}5\\7\end{bmatrix};\ 28 : \begin{bmatrix}6\\7\end{bmatrix};$$

A set of blocks S_2 obtained by trial is given below:

$8 : \begin{bmatrix}3\\7\end{bmatrix};\ 9 : \begin{bmatrix}4\\5\end{bmatrix};\ 10 : \begin{bmatrix}2\\6\end{bmatrix};\ 11 : \begin{bmatrix}3\\6\end{bmatrix};\ 12 : \begin{bmatrix}4\\7\end{bmatrix};\ 13 : \begin{bmatrix}2\\5\end{bmatrix};\ 14 : \begin{bmatrix}5\\6\end{bmatrix};$

$15 : \begin{bmatrix}5\\7\end{bmatrix};\ 16 : \begin{bmatrix}1\\4\end{bmatrix};\ 17 : \begin{bmatrix}3\\4\end{bmatrix};\ 18 : \begin{bmatrix}1\\6\end{bmatrix};\ 19 : \begin{bmatrix}6\\7\end{bmatrix};\ 20 : \begin{bmatrix}1\\7\end{bmatrix};\ 21 : \begin{bmatrix}1\\2\end{bmatrix};$

$22 : \begin{bmatrix}2\\4\end{bmatrix};\ 23 : \begin{bmatrix}2\\3\end{bmatrix};\ 24 : \begin{bmatrix}1\\5\end{bmatrix};\ 25 : \begin{bmatrix}1\\3\end{bmatrix};\ 26 : \begin{bmatrix}2\\7\end{bmatrix};\ 27 : \begin{bmatrix}4\\6\end{bmatrix};\ 28 : \begin{bmatrix}3\\5\end{bmatrix}$

The following set of β-chains yield a solution which is found to be different from the solutions obtained by the "Process of block section" of the classical solution given by Bose (1939), Fisher and Yates (1938, 1942) and of the solutions given by Husain (1945).

(B).

$8 : \begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}3\\7\end{bmatrix}$ (12)(37)(4564) $9 : \begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}4\\5\end{bmatrix}$ (13)(45)(2672) $10 : \begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}2\\6\end{bmatrix}$ (14)(26)(3573)

$11 : \begin{bmatrix}1\\5\end{bmatrix}\begin{bmatrix}3\\6\end{bmatrix}$ (15)(36)(2742) $12 : \begin{bmatrix}1\\6\end{bmatrix}\begin{bmatrix}4\\7\end{bmatrix}$ (16)(47)(2352) $13 : \begin{bmatrix}1\\7\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}$ (17)(25)(3463)

$14 : \begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}5\\6\end{bmatrix}$ (23)(56)(1471) $15 : \begin{bmatrix}2\\4\end{bmatrix}\begin{bmatrix}5\\7\end{bmatrix}$ (24)(57)(1361) $16 : \begin{bmatrix}2\\5\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$ (25)(14)(3673)

$17 : \begin{bmatrix}2\\6\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$ (26)(34)(1571) $18 : \begin{bmatrix}2\\7\end{bmatrix}\begin{bmatrix}1\\6\end{bmatrix}$ (27)(16)(3453) $19 : \begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}6\\7\end{bmatrix}$ (34)(67)(1251)

$20 : \begin{bmatrix}3\\5\end{bmatrix}\begin{bmatrix}1\\7\end{bmatrix}$ (35)(17)(2462) $21 : \begin{bmatrix}3\\6\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$ (36)(12)(4574) $22 : \begin{bmatrix}3\\7\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$ (37)(24)(1651)

$23 : \begin{bmatrix}4\\5\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}$ (23)(45)(1671) $24 : \begin{bmatrix}4\\6\end{bmatrix}\begin{bmatrix}1\\5\end{bmatrix}$ (15)(46)(2372) $25 : \begin{bmatrix}4\\7\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}$ (13)(47)(2562)

$26 : \begin{bmatrix}5\\6\end{bmatrix}\begin{bmatrix}2\\7\end{bmatrix}$ (56)(27)(1341) $27 : \begin{bmatrix}5\\7\end{bmatrix}\begin{bmatrix}4\\6\end{bmatrix}$ (57)(46)(1231) $28 : \begin{bmatrix}6\\7\end{bmatrix}\begin{bmatrix}3\\5\end{bmatrix}$ (67)(35)(1241)

Let (B_1) be the design obtained by deleting the second set of links from each of the above β-chains and similarly (B_2) be the design obtained by deleting the first set of links from each of the β-chains. The two designs so obtained are isomorphic, for the transposition of initial varieties, (1, 2), (4,5), and (6, 7) converts the whole set of π-chains of (B_2) into those of (B_1) and when the ψ-varieties corresponding to the identified chains are identified, the design (B_2) becomes identical with (B_1).

It is interesting to note that the dual of a symmetrical PBIB design, gives not necessarily the same solution of the design, though the dual of a dual is the original design. This is brought out by the example of the dual of (B_1) as recorded below:

(B_1').

$8 : \begin{bmatrix}6\\7\end{bmatrix}$ (67)(13452) ; $9 : \begin{bmatrix}5\\7\end{bmatrix}$ (57)(14263)

$10 : \begin{bmatrix}5\\6\end{bmatrix}$ (56)(12374) ; $11 : \begin{bmatrix}3\\4\end{bmatrix}$ (34)(17265)

$12 : \begin{bmatrix}2\\4\end{bmatrix}$ (24)(15376) ; $13 : \begin{bmatrix}2\\3\end{bmatrix}$ (23)(16457)

$14 : \begin{bmatrix}4\\6\end{bmatrix}$ (46)(21753) ; $15 : \begin{bmatrix}3\\5\end{bmatrix}$ (35)(27614)

$16 : \begin{bmatrix}4\\7\end{bmatrix}$ (47)(23615) ; $17 : \begin{bmatrix}1\\3\end{bmatrix}$ (13)(24756)

$18 : \begin{bmatrix}1\\5\end{bmatrix}$ (15)(26437) ; $19 : \begin{bmatrix}2\\7\end{bmatrix}$ (27)(31564)

$20 : \begin{bmatrix}1\\6\end{bmatrix}$ (16)(37245) ; $21 : \begin{bmatrix}2\\5\end{bmatrix}$ (25)(34716)

$22 : \begin{bmatrix}1\\4\end{bmatrix}$ (14)(32567) ; $23 : \begin{bmatrix}1\\2\end{bmatrix}$ (12)(43675)

$24 : \begin{bmatrix}1\\7\end{bmatrix}$ (17)(45326) ; $25 : \begin{bmatrix}3\\6\end{bmatrix}$ (36)(42517)

$26 : \begin{bmatrix}3\\7\end{bmatrix}$ (37)(52146) ; $27 : \begin{bmatrix}2\\6\end{bmatrix}$ (26)(53147)

$28 : \begin{bmatrix}4\\5\end{bmatrix}$ (45)(63127).

It appears that all the rings in the π-chains of (B_1') are formed of one simple ring and another ring of five elements. Therefore (B_1') is non-isomorphic with (B_1).

The classical solution of the symmetrical BIB design for $n = 8$ and $\lambda = 2$ is obtained, as is well-known, by starting with the initial block (1, 16, 34, 26, 9, 33, 10, 12, 7), and by forming from it the 36 blocks derivable by the addition of the numbers 1, 2, ..., 36, taking residues (mod 37). By omitting the initial block and its varieties from the other blocks and then substituting the variety (17, 1) and thereafter interchanging the pairs of varieties, (35, 3); (27, 4); (11, 5); (13, 6); (8, 7); as well as renaming the varieties, (11, 28, 23, 18, 27, 29, 30, 32, 36, 37, 13, 14, 15, 19, 20, 21, 22, 24, 25, 31, 35) by the respective varieties 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, we get the following BIB design.

(*C*).

8 : $\begin{bmatrix}1\\7\end{bmatrix}\begin{bmatrix}3\\7\end{bmatrix}$ (165243) ;

9 : $\begin{bmatrix}2\\4\end{bmatrix}\begin{bmatrix}1\\4\end{bmatrix}$ (167352)

10 : $\begin{bmatrix}2\\5\end{bmatrix}\begin{bmatrix}2\\5\end{bmatrix}$ (163471) ;

11 : $\begin{bmatrix}1\\5\end{bmatrix}\begin{bmatrix}1\\3\end{bmatrix}$ (26472)(35)

12 : $\begin{bmatrix}1\\3\end{bmatrix}\begin{bmatrix}4\\6\end{bmatrix}$ (17254)(36) ;

13 : $\begin{bmatrix}5\\7\end{bmatrix}\begin{bmatrix}2\\6\end{bmatrix}$ (64137)(25)

14 : $\begin{bmatrix}4\\7\end{bmatrix}\begin{bmatrix}1\\5\end{bmatrix}$ (56327)(14) ;

15 : $\begin{bmatrix}3\\5\end{bmatrix}\begin{bmatrix}4\\7\end{bmatrix}$ (32614)(57)

16 : $\begin{bmatrix}2\\6\end{bmatrix}\begin{bmatrix}6\\7\end{bmatrix}$ (14351)(27) ;

17 : $\begin{bmatrix}2\\3\end{bmatrix}\begin{bmatrix}1\\6\end{bmatrix}$ (24756)(13)

18 : $\begin{bmatrix}2\\7\end{bmatrix}\begin{bmatrix}3\\4\end{bmatrix}$ (2157)(364) ;

19 : $\begin{bmatrix}3\\4\end{bmatrix}\begin{bmatrix}2\\7\end{bmatrix}$ (3564)(217)

20 : $\begin{bmatrix}1\\4\end{bmatrix}\begin{bmatrix}5\\6\end{bmatrix}$ (4375)(126) ;

21 : $\begin{bmatrix}4\\5\end{bmatrix}\begin{bmatrix}3\\6\end{bmatrix}$ (5176)(324)

22 : $\begin{bmatrix}1\\6\end{bmatrix}\begin{bmatrix}2\\4\end{bmatrix}$ (2374)(156) ;

23 : $\begin{bmatrix}3\\6\end{bmatrix}\begin{bmatrix}3\\5\end{bmatrix}$ (521476)

24 : $\begin{bmatrix}3\\7\end{bmatrix}\begin{bmatrix}1\\2\end{bmatrix}$ (1543)(267) ;

25 : $\begin{bmatrix}1\\2\end{bmatrix}\begin{bmatrix}5\\7\end{bmatrix}$ (5467)(132)

26 : $\begin{bmatrix}4\\6\end{bmatrix}\begin{bmatrix}2\\3\end{bmatrix}$ (2754)(316) ;

27 : $\begin{bmatrix}5\\6\end{bmatrix}\begin{bmatrix}1\\7\end{bmatrix}$ (1245)(637)

28 : $\begin{bmatrix}6\\7\end{bmatrix}\begin{bmatrix}4\\5\end{bmatrix}$ (5317)(426).

Similarly, the two BIB designs derived from the solutions given by Husain (1945) and when expressed in our chains, are seen to be different from those (*B*) and (*C*).

Let (C_1) be the π-design obtained by deleting the second set of links from the chains of the above design (C), and (C_2) be the design obtained by deleting the first set of links. Clearly, (C_1) and (C_2) are non-isomorphic for the π-chain corresponding to the variety 10 in (C_1) cannot be identified with any other chain than that of variety 10 in (C_2) which is the only chain of a similar type. Any interchange of the elements of the other similar type of chains in either of the designs will disturb the already identified chain for the variety 10. Thus we get two independent solutions of the π-design from the classical design. The totality of independent solutions for π with $n = 8$ awaits study.

4. EQUIVALENCE TEST OF TWO METHODS FOR DERIVING CERTAIN DESIGN FROM A KNOWN SYMMETRICAL BIB DESIGN WITH $\lambda = 2$

Shrikhande (1952) showed that by omitting all the blocks containing a particular variety from a known symmetrical BIB design with $r = n+1$ and $\lambda = 2$, a PBIB design with the following parameters is obtained :

$$v = \frac{n(n+1)}{2}, \quad b = \frac{n(n-1)}{2}, \quad k = n+1, \quad r = n-1, \quad \lambda_1 = 2, \quad \lambda_2 = 1,$$

$$n_1 = \frac{(n-1)(n-2)}{2}, \quad n_2 = 2n-2,$$

$$P^1_{jk} = \begin{pmatrix} \frac{(n-3)(n-4)}{2} & 2n-6 \\ 2n-6 & 4 \end{pmatrix},$$

$$P^2_{jk} = \begin{pmatrix} \frac{(n-2)(n-3)}{2} & n-2 \\ n-2 & n-1 \end{pmatrix}. \qquad \ldots \quad (4.1)$$

It is also well-known that by first omitting any block and its varieties from the other blocks of the same symmetrical BIB design with $\lambda = 2$ and then by dualising the residual configuration, a PBIB design satisfying the above parameters is obtained. The two ways of deriving such PBIB designs from the same existing symmetrical BIB design may be called shortly M_1 and M_2 and the two solutions so obtained P_1 and P_2 respectively. Our main object it to study if the two methods would always lead to isomorphic solutions.

We can select groups of n varieties, X_j $(j = 1, \ldots n)$ which are mutually second associates in the PBIB design (4.1), any one of these groups having the property that every pair of its varieties is represented uniquely in the different blocks of the design, so that the blocks may be specified uniquely by the different pairs of the X-varieties. The location of a Y-variety in the $r = (n-1)$ blocks can therefore be expressed by naming the X-variety pairs occurring in the blocks containing the particular Y-variety. We can define the $(n-1)$ blocks containing a particular Y-variety by means of a 'chain', taken to be a P-chain, which may contain one or more rings. Since a Y-variety has $N_1 = \frac{(n-2)(n-3)}{2}$ first associates and $N_2 = 2n-4$, second associates among the remaining Y-varieties, it implies that any particular chain should have exactly 2 blocks common with $\frac{(n-2)(n-3)}{2}$ chains and 1 block common with $2n-4$ remaining chains, thus accounting for all the Y-varieties.

A set of P-chains may be called 'consistent', if any particular chain shall have exactly two pairs of consecutive numbers common with $\frac{(n-2)(n-3)}{2}$ other chains; and exactly one pair of consecutive numbers common with the remaining $2n-4$ chains. It is thus clear that if there be given a PBIB design (4.1) there will correspond to each Y-variety one distinct chain of a consistent set.

Consider the symmetrical BIB design with $r = 5$ and $\lambda = 2$. The design is obtained, as is well-known, by adding to the elements of an initial block (1, 4, 5, 9, 3) the numbers 1, 2, ..., 10 and taking residues (mod 11). Let U : (9, 1, 2, 6, 11) be a block of this design. Then we get by the method M_1 (i.e. by ignoring the blocks containing the variety 11) the following solution which is expressed conveniently in terms of the P-chains of its Y-varieties with reference to the X-elements 1, 2, 6 and 9.

3: (2 1 9 6)   4: (2 6 1 9)
P_1 5: (1 9 2 6)   7: (9 2 1 6)
8: (1 6 9 2)   10: (1 2 6 9)

Again, by the method M_2 (i.e. first by omitting the block U and its varieties, viz., 9, 1, 2, 6, 11 from the other blocks, and then dualising the residual configuration) we get the following design :

5: (2 4 1 3)   6: (1 3 2 4)
P_2 7: (1 2 4 3)   8: (2 1 3 4)
9: (1 4 3 2)   10: (3 2 1 4)

It is easy to see that the transposition of the varieties (3,9) and (4,6) charges the chains of P_1 into those of P_2 and the Y-varieties corresponding to the identified chains then being identified make the two solutions identical. The other solutions obtained either by trial or by the methods M_1 and M_2 with respect to different varieties have been found to be associated with similar type of chains, each formed of a single ring of four elements, and all of them, on interchanges of suitable varieties, become identical, thereby showing that in this case the two methods lead to the identical solution.

Husain (1945) demonstrated that the symmetrical BIB design with $r = 6$ and $\lambda = 2$ admits of three independent solutions including the classical solution given by Bose (1939), and Fisher and Yates (1938) which are isomorphic. The operations M_1 and M_2 when applied to any one of the three independent solutions give isomorphic solutions for the PBIB design (4.1). For example, if we name the sixteen alphabets in Fisher's design (1938) from 1 to 16 and if we omit all the blocks containing the variety 6 of the block U : (1, 2, 3, 4, 5, 6), then the chains of the Y-varieties for the solutions P_1 and P_2 derived by the methods M_1 and M_2 are respectively given by:

P_1.

7: (1231)(45) 8: (1241)(35)
9: (1251)(34) 10: (3453)(12)
11: (1341)(25) 12: (1351)(24)
13: (2452)(13) 14: (1451)(23)
15: (2352)(14) 16: (2342)(15)

P_2.

6: (3453)(12) 7: (2452)(13)
8: (2352)(14) 9: (2342)(15)
10: (1451)(23) 11: (1351)(24)
12: (1341)(25) 13: (1251)(34)
14: (1241)(35) 15: (1231)(45).

Clearly, the set of chains in P_1 and P_2 are equivalent and therefore the two solutions on proper identification of their Y-varieties become isomorphic. Similarly, a solution derived by M_1 or by M_2 from either of the other two independent solutions of the above symmetrical BIB design are found to be associated with six chains of the type (12345) and four chains of the type (12)(3453) in one case, and six chains of the type (12)(3453) and four chains of the type (12345) in the other.

There are altogether seven independent solutions known for the symmetrical BIB design with $r = 9$ and $\lambda = 2$, one is the classical solution given by Bose (1939), Fisher and Yates (1938), two are due to Husain (1945), and four are due to the author discussed elsewhere (1956). The methods M_1 and M_2 when applied to these independent solutions do not necessarily lead to the isomorphic solutions. For example, consider the following symmetrical BIB design which is obtained by the author elsewhere (1956) and is presented in terms of Husain's (1945) λ-chains.[1]

(E).

$\overline{1}$: (123)(478)(569)
$\overline{2}$: (128397654)
$\overline{3}$: (126849735)
$\overline{4}$: (129475836)
$\overline{5}$: (124358967)
$\overline{6}$: (127364598)
$\overline{7}$: (125786439)
$\overline{8}$: (137568924)
$\overline{9}$: (139276485)
$\overline{10}$: (135289746)
$\overline{11}$: (138694527)
$\overline{12}$: (134957628)
$\overline{13}$: (136245879)
$\overline{14}$: (149876325)
$\overline{15}$: (143279586)
$\overline{16}$: (146295387)
$\overline{17}$: (147396258)
$\overline{18}$: (148275369)
$\overline{19}$: (154378296)
$\overline{20}$: (159368247)
$\overline{21}$: (157964238)
$\overline{22}$: (156274389)
$\overline{23}$: (162398457)
$\overline{24}$: (165394278)
$\overline{25}$: (167384259)
$\overline{26}$: (179256348)
$\overline{27}$: (173285649)
$\overline{28}$: (186745329)

---

[1] A λ-chain consists of one or more cycles. A cycle must contain 3 or more elements of a standard block, usually called an initial block of a symmetrical BIB design with $\lambda = 2$. It can begin at any element and can be read either forward or backward, the element at the beginning and at the end being regarded as consecutive. Three consecutive numbers of one chain must not coincide with three numbers of the other chain (read forward and backward). Exactly two pairs of consecutive numbers should be common between any two chains.

The above symmetrical design when subjected to the operations M_1 and M_2 with reference to the block U: (1, 2, 3, ..., 9) and the other blocks containing the variety 1 gives the following designs P_1 and P_2 respectively.

P_1	$\bar{1}$: (23)(4784)(5695)	$\bar{2}$: (45679382)	P_2	$\bar{1}$: (2342)(59786)	$\bar{2}$: (2352)(49678)
	$\bar{3}$: (26849735)	$\bar{4}$: (29475836)		$\bar{3}$: (2362)(48957)	$\bar{4}$: (2372)(65849)
	$\bar{5}$: (24358967)	$\bar{6}$: (27364598)		$\bar{5}$: (2382)(57469)	$\bar{6}$: (2392)(74568)
	$\bar{7}$: (25786439)	$\bar{8}$: (37568924)		$\bar{7}$: (23)(4589764)	$\bar{8}$: (26534789)
	9: (39276485)	$\overline{10}$: (35289746)		$\bar{9}$: (82763459)	$\overline{10}$: (64375298)
	$\overline{11}$: (38694527)	$\overline{12}$: (34957628)		$\overline{11}$: (76294385)	$\overline{12}$: (27586934)
	$\overline{13}$: (36245879)	$\overline{14}$: (49876325)		$\overline{13}$: (34825697)	$\overline{14}$: (47928536)
	$\overline{15}$: (43279586)	$\overline{16}$: (46295387)		$\overline{15}$: (28649537)	$\overline{16}$: (69835427)
	$\overline{17}$: (47396258)	$\overline{18}$: (48275369)		$\overline{17}$: (53972648)	$\overline{18}$: (35768429)
	$\overline{19}$: (54378296)	$\overline{20}$: (59368247)		$\overline{19}$: (29637845)	$\overline{20}$: (24975638)
	$\overline{21}$: (57964238)	$\overline{22}$: (56274389)		$\overline{21}$: (78524639)	$\overline{22}$: (36894725)
	$\overline{23}$: (62398457)	$\overline{24}$: (65394278)		$\overline{23}$: (45268379)	$\overline{24}$: (42893765)
	$\overline{25}$: (67384259)	$\overline{26}$: (79256348)		$\overline{25}$: (37426958)	$\overline{26}$: (25938476)
	$\overline{27}$: (73285649)	$\overline{28}$: (86745329)		$\overline{27}$: (38729564)	$\overline{28}$: (39457826)

It appears that in P_1 there are twentyseven P-chains of the type (23456789) and just one chain of the type (23)(4564)(7897) while in P_2 there are twentyone chains of the type (23456789); six chains of the type (2342) (56789) and just one chain of the type (23)(456789). Hence the two solutions are not isomorphic. But the classical solution with the initial block U: (1, 7, 9, 10, 12, 16, 26, 33, 34) of residues (mod 37), when subjected to the operations M_1 and M_2 with reference to the set of blocks containing the variety 34, yields two solutions, each having its Y-varieties associated with nineteen chains of the type (23456789); five chains of the type (23452)(6789) and four chains of the type (234562) (789), and both the solutions come out to be isomorphic by suitable substitutions. Thus it follows that the operations M_1 and M_2 do not give in general identical solutions and consequently non-isomorphic solutions for the BIB design W or for the PBIB design (4.1) may be obtained from the same solution of a known symmetrical BIB design with $r = n+1$ and $\lambda = 2$.

I wish to express my thanks to Professor Q. M. Husain for his encouragement and helpful suggestions. I would like to thank the referees for their comments which led to the improvement of the paper.

REFERENCES

ATIQULLAH, M. (1958): Some new solutions of the symmetrical balanced incomplete block design with $k = 9$ and $\lambda = 2$. *Bull. Cal. Math. Soc.* **50**, 23.

BOSE, R. C. (1939): On the construction of balanced incomplete block design. *Ann. Eugen*, **9**, 353.

BOSE, R. C. AND SHIMAMOTO, T. (1952): Classification and analysis of partially balanced incomplete block design with two associate classes. *Amer. Stat. Ass.*, **47**, 151.

BOSE, R. C. AND NAIR, K. R. (1939): Partially balanced incomplete block design. *Sankhyā*, **4**, 337.

COX, GERTUDE, M. (1940): Enumeration and construction of balanced incomplete configurations. *Ann. Math. Stat.*, **9**, 72.

CONNOR, W. S. (1952): On the structure of balanced incomplete block design. *Ann. Math. Stat.*, **23**, 57.

CONNOR, W. S. AND CLATWORTHY WH. (1954): Some theorems of partially balanced designs. *Ann. Math. Stat.*, **25**, 100.

FISHER, R. A. (1940): An examination of the different possible solutions of a problem in incomplete block. *Ann. Eugen.*, **10**, 52.

FISHER, R. A. AND YATES, F. (1936, 1942): *Statistical Tables for Biological, Agricultural and Medical Research*, Oliver and Boyd, London.

HALL, MARSHAL AND CONNOR, W. S. (1953): An embodying theorem for balanced incomplete block designs. *Cand. J. of Math.*, **6**, 35.

HUSAIN, Q. M. (1945): Symmetrical incomplete block design with $\lambda = 2$, $k = 8$ or 9. *Bull. Cal. Math. Soc.*, **37**, 116.

——— (1945a): On the totality of the solution of symmetrical incomplete block design with $\lambda = 2$, $k = 5$ or 6. *Sankhyā*, **7**, 204.

——— (1948): Structure of some incomplete block designs. *Sankhyā*, **8**, 381.

NANDI, H. K. (1945): Enumeration of non-isomorphic solutions of balanced incomplete block designs. *Sankhyā*, **7**, 305.

——— (1946a): Enumeration of non-isomorphic solutions of balanced incomplete block designs. *Sankhyā*, **7**, 305.

——— (1946b): A further note on non-isomorphic solutions of incomplete block designs. *Sankhyā*, **7**, 313.

NAIR, K. R. AND RAO, C. R. (1942): A note on partially balanced incomplete block designs. *Science and Culture*, **7**, 568.

NAIR, K. R. (1951): Rectangular lattices and partially balanced incomplete block designs. *Biometrics*, **7**, 145.

SHRIKHANDE, S. S. (1950): The Impossibility of certain symmetrical balanced incomplete block designs. *Ann. Math. Stat.*, **21**, 106.

——— (1952): On the dual of some balanced incomplete block design. *Biometrics*, **8**, 66.

Paper received : November, 1956.

Revised : May, 1957.

TABLES OF RANDOM NORMAL DEVIATES

By J. M. SENGUPTA
and
NIKHILESH BHATTACHARYA
Indian Statistical Institute, Calcutta

FOREWORD

The present series was published in 1934 in Sankhyā, Vol. 1, pp. 289–328, with an introduction explaining its uses and especially describing the results of elaborate statistical tests carried out on the series. These tests showed that the series was of satisfactory quality, which was, of course, somewhat equivalent to confirming the quality of Tippett's random numbers. Since then a number of persons, including some workers of the Indian Statistical Institute, have detected errors in the conversion work.[1]

When we first became aware of the discrepancies we found that certain portions of the table had been changed in quite an arbitrary manner. In the early days of Sankhyā, we did not have a printing press of our own. We had purchased some mathematical and other types and we arranged for the matter to be composed in my own residence. The composed pages were then placed in flat boxes with closed covers and were carried to the Art Press where the printing was done. It is possible that some of the composed matter, but not all, had been accidentally disturbed at the last stage and the men on the spot had re-arranged the matter before printing it. We believe this is a plausible explanation.

Revised tables were accordingly prepared in the Institute in 1952. In 1954, it was learnt that Miss Lorrain Schwartz had independently prepared a set of the tables in the Statistical Laboratory at Berkeley, California. A copy of the revised table was then sent to Berkeley, and Professor Henry Scheffé in his letter of 4 June 1955 informed us that the two tables checked perfectly.

The revised table is now being published along with a fresh introduction. It has been considered advisable to re-examine the series. The tests applied are broadly similar to those applied on the other series of the same type, but naturally there are some differences in respect of details.

5 April, 1958 P. C. MAHALANOBIS

---

[1] Mr. Dan Teichroew, in a letter from North Carolina dated 19 May 1952 pointed out some discrepancies between the *Sankhyā* series and a short table by E. L. Dodd (based on Tippett's *Tract*) published in *Boletin Matematico*, Vol. 15, 1942.

TABLES OF RANDOM NORMAL DEVIATES

By J. M. SENGUPTA
and
NIKHILESH BHATTACHARYA
Indian Statistical Institute, Calcutta

1. GENERAL INTRODUCTION

1.1. The advantages of having tables of random sampling numbers were first realised by Professor Karl Pearson, at whose instance L. H. C. Tippett compiled a set of 41,600 random digits about 1925 and published them as *Tracts for Computers*, no. XV, in 1927. Since then different series of such numbers have appeared, culminating in the publication by the RAND Corporation in 1955 of a series of 'a million random digits.'

1.2. Besides making possible for the first time the rapid and rigorous selection of (even large) samples from diverse kinds of existent populations, these numbers lent themselves to a variety of uses covered by the now familiar term "model sampling". Briefly, model sampling provides an experimental method for investigating sampling distributions of various statistics, and can therefore be used for demonstrating known results of sampling theory.

1.3. As much of statistical theory centres round the normal distribution, model sampling was soon found to involve, very frequently, the conversion of random numbers into random normal deviates, a process tedious for any large-scale work. This naturally suggested the systematic conversion of tables for random sampling numbers to derive corresponding tables for random normal deviates. The work was actually carried out with Tippett's series by P. C. Mahalanobis and others of the Indian Statistical Institute, and the resulting table of 10,400 deviates appeared in 1934, in Sankhyā, Vol. 1, pp. 289–328. The need for more extensive tables was felt by Wold in connection with studies in time series analysis. Wold (1948) converted Kendall and Babington Smith's (1939) table of 100,000 random digits into a table of 25,000 random normal deviates issued as *Tracts for Computers*, no. XXV. Still more extensive tables were required in connection with problems solved by what have come to be known as Monte Carlo methods, and recently the RAND Corporation has published a series of 100,000 such deviates derived from half of their 'a million random digits'.

1.4. Artificial random samples from a normal population are frequently used for the experimental confirmation and/or illustration of known *normal theory* results. They are also of use in obtaining information on theoretically or numerically intractable sampling distributions of statistics. As an important sub-class of this latter type of situations we have the case where only the limiting distribution (as 'sample size' n tends to infinity) is known, but one wants to know how far, and for what value of n onwards, this limiting distribution can be trusted. As emphasised by

Wold, such experimental work on sampling distributions generally requires longer and more carefully constructed series of random normal deviates.

1.5. Classic examples of the use of model sampling experiments are provided by "Student's" pioneer work (Pearson and Wishart, 1942). (In the absence of random sampling numbers, Student used card sampling, cards corresponding to individuals of the existent population.) His paper on "The probable error of a mean" contained experimental confirmation of the theoretical results derived, the artificial samples being taken from an actual (bivariate) population very nearly normal. The same method was adopted in his paper entitled "Probable error of a correlation coefficient" to obtain preliminary insight into the exact distribution of sample correlation coefficient r in small samples for the case where the population correlation coefficient $\rho = 0$; and again, in a similar manner, for studying the "Probable error of Dr. Spearman's correlation coefficient."

1.6. References (Hojo, 1931; Pepper, 1932; Pearson, 1932; Rider, 1932) cited at the end of this introduction contain further examples of model sampling experiments. In the absence of tables of random normal deviates, these investigations were done on the basis of Tippett's numbers.

1.7. Tables of random normal deviates are frequently used for constructing artificial time series for model sampling experiments. As examples one may cite the experimental studies on the applicability of large sample tests to time series of short length, and on the estimation of parameters in a recursive system (Matthai and Kanan, 1951; Rao and Som, 1951; Das, 1951).

2. CONSTRUCTION AND USES

2.1. The present tables have been constructed by the method indicated in (Illustration I, p. iv of) K. Pearson's foreword to Tippett's tables. Refinements of the type described in the introduction to Wold's series have not been adopted. However, this omission does not seem to have affected the quality of the series appreciably, as shown by the satisfactory results of tests of significance reported below, based on distributions of range for sample sizes $n = 10, 11, 12$, and 15.

2.2. The tables were constructed by punched card methods. Tippett's four-digited random numbers were punched into one set of cards, and the corresponding normal deviates were read off from The Kelley Statistical Tables (1948) transferred to another set of punched cards.[1]

2.3. Any entry in the present table is a random sample from a normal population having zero mean and unit s.d. If x be the entry, then $m+\sigma x$ $(\sigma \geqslant 0)$ gives a random sample from a normal population having mean m and s.d. σ.[1]

---

[1] Instructions to Machine Tabulation Section were as follows:

"...This may be done by using Kelley's Table (already punched in cards) and *collating* the four-digited random numbers against Kelley's p-values, when they (random numbers) are 5,000 or above, but against $(1-p)$-values when they are less than 5,000. ...The normal deviates corresponding to random numbers equal to or exceeding 5,000 will be positive, while those corresponding to random numbers below 5,000 will be negative."

2.4. In order to form samples of any given size one may take the required number of entries vertically, horizontally or in any other way. Strictly, however, one should read the table vertically, exhausting one column after another, (this being the manner of using the table in carrying out the statistical tests,) although it is very unlikely that reading in other directions would lead to biased samples. In any case, it is not advisable to multiply the sample size by reading off the tables in various directions, for this might give rise to unwanted correlations. For large-scale work requiring larger number of deviates, one should use the other two tables.

2.5. The common procedure of 'opening at a page haphazardly etc.' may be refined as follows. First, select one plate number i at random from among plate numbers 1 to 26, then a column number j from among numbers 1 to 8, and lastly, one row number k from among 1 to 50. Then start at the k-th deviate from top along the j-th column of the i-th plate, and read vertically, exhausting one column before going to the top of the next.

2.6. The usual procedure of rounding can be followed when normal deviates with fewer places of decimals are wanted. As regards figures ending with a 5 (or 50), one might raise or keep unchanged the last figure to be retained according as the next deviate has the sign '+' or '—'.

2.7. A table of random normal deviates can be used for sampling from theoretical distributions related to the normal distribution, which between themselves cover a wide area of the Pearsonian family of curves. This indirect method may suffice for most investigations in non-normal theory.

2.8. Another use of random normal deviates has been described in the introduction to Wold's table. Suppose one requires to draw a random sample from a four-variate normal population, with means 2.5, –17.2, 24.0 and 140.0 respectively, and a dispersion matrix

$$\Lambda = \begin{pmatrix} 32.95 & 7.43 & 1.78 & 3.97 \\ 7.43 & 10.24 & 1.17 & 2.43 \\ 1.78 & 1.17 & 3.06 & 1.78 \\ 3.97 & 2.43 & 1.78 & 12.25 \end{pmatrix}.$$

The first step is to find a triangular matrix of the form

$$P = \begin{pmatrix} p_{11} & \text{—} & \text{—} & \text{—} \\ p_{21} & p_{22} & \text{—} & \text{—} \\ p_{31} & p_{32} & p_{33} & \text{—} \\ p_{41} & p_{42} & p_{43} & p_{44} \end{pmatrix}$$

such that $PP' = \Lambda$, where P' is the transpose of P. The solutions come out as follows:

$$P = \begin{pmatrix} 5.740209 & \text{—} & \text{—} & \text{—} \\ 1.294378 & 2.926531 & \text{—} & \text{—} \\ 0.310093 & 0.262639 & 1.701430 & \text{—} \\ 0.691612 & 0.524441 & 0.839175 & 3.285182 \end{pmatrix}.$$

---

[1] Throughout the table, '+' and '—' signs of deviates have been indicated by codes '1' and '2' respectively. (Vide footnotes on any plate.)

One has then to read four values of random normal deviates denoted by say, x_1, x_2, x_3 and x_4. Suppose $x_1 = -1.194$, $x_2 = -1.406$, $x_3 = 0.868$, and $x_4 = 0.076$. If now one calculates

$$\begin{aligned} y_1 &= p_{11}x_1+2.5 = -4.354,\\ y_2 &= p_{21}x_1+p_{22}x_2-17.2 = -22.860,\\ y_3 &= p_{31}x_1+p_{32}x_2+p_{33}x_3+24.0 = 24.737,\\ y_4 &= p_{41}x_1+p_{42}x_2+p_{43}x_3+p_{44}x_4+140.0 = 138.916, \end{aligned}$$

then (y_1, y_2, y_3, y_4) is a random sample from the multi-normal population specified above.

2.9. As mentioned in Karl Pearson's foreword to Tippett's *Tracts*, and later, more explicitly, in the introduction to Kendall and Babington Smith's *Tracts*, all the small sub-samples of random digits or random normal deviates in such tables are not expected to pass any or all statistical test or tests. Rather if the construction of the tables has been alright, a proportion of sub-samples (indicated by the level of significance) should fail to satisfy the test. These patches are not suitable for use by themselves but should in all fairness be included when large number of deviates are being utilised.

2.10. Sub-samples failing to satisfy the different tests of (local) quality at the 1% level of significance are listed below. Notes are also being given on the body of the tables at appropriate places.

1. Plate nos. 7 (whole) and 8 (cols. 1 to 4) : failed in tests based on range with sample size $n=6$ (see below).
2. Plate no. 17 (whole) : —do— with $n=2$.

3. THE NORMALITY OF THE SAMPLE

3.1. The value of the present table depends on how far the deviates are independent random samples from a standard normal population. This point has been tested in two different ways, both of which involved heavy computational labour. The first was to compare the frequency distribution of these deviates with the theoretical distribution, for the whole sample and its various sub-samples. The other was to use these tables for model sampling experiments on the distribution of range (for various sizes of samples) and then to compare the distributions of range thus obtained with the known normal theory distributions tabulated by Pearson and Hartley (1942). These two series of tests are being reported in this and in the following sections.

3.2. In preparing frequency distributions of the deviates the intervals used were of length 0.2, i.e. one-fifth of the standard deviation, which is, of course, quite adequate. (The intervals were: ..., −0.2 to 0, 0 to 0.2, 0.2 to 0.4, etc.) Separate frequency distributions were prepared for each of the 104 exclusive sets of 100 entries occurring in two consecutive columns in the tables. The χ^2–test of goodness of fit was applied to each of these 104 distributions. When pooling class-intervals so as to avoid very small expected frequencies, Cochran's (1952) recommendations were followed throughout the present investigation. For the χ^2–tests being reported in paragraphs 3.2 to 3.5, the 'minimum expectation' was taken to be 1 at both the tails. Thus for frequency distributions of 100 deviates, the pooled class-intervals were : $-\infty$ to -2.0, -2.0 to -1.8, …, 1.8 to 2.0, 2.0 to $+\infty$. Fourteen of the χ^2-values were significant at the 10% level, out of which 4 exceeded the 5% level but none exceeded the 1% level. The sum of these 104 χ^2's came out as 2174.0, which is obviously non-significant, since the degrees of freedom corresponding to this figure, which is a χ^2, is 2184.

3.3. Table 1 shows the frequency distribution of these 104 values of χ^2, together with a comparison with the normal theory distribution of χ^2 with 21 degrees of freedom. The value of χ^2 with 15 degrees of freedom was 13.840.

TABLE 1. OBSERVED AND EXPECTED FREQUENCY DISTRIBUTIONS OF 104 GOODNESS OF FIT χ^2's WITH 21 d.f.

class-interval	observed frequency (f_o)	expected frequency (f_e)
(1)	(2)	(3)
0.0– 2.0	0	0.002
2.0– 4.0		
4.0– 6.0	0	0.057
6.0– 8.0	0	0.446
8.0–10.0	5	1.688
10.0–12.0	2	4.086
12.0–14.0	7	7.282
14.0–16.0	8	10.395
16.0–18.0	15	12.548
18.0–20.0	14	13.285
20.0–22.0	12	12.662
22.0–24.0	11	11.071
24.0–26.0	8	9.007
26.0–28.0	5	6.895
28.0–30.0	4	5.009
30.0–32.0	7	3.478
32.0–34.0	4	2.320
34.0–36.0	2	1.494
36.0–39.0	0	1.254
39.0–∞	0	1.020
total	104	104.000

χ^2 with 15 d.f. = 13.840.

3.4. The frequency distributions for all the 400 entries in each of the 26 plates were then obtained by combining the four distributions, each for 100 entries, prepared from the same page of the tables. The χ^2-test was applied to each of these distributions. The class-intervals after pooling were : $-\infty$ to -2.6, -2.6 to -2.4, …, 2.4 to 2.6, 2.6 to $+\infty$. Three of these 26 χ^2's were significant at the 10% probability level, out of which two were significant at the 5% level, but none at the 1% level. The sum of these values is only 706.78, which is a χ^2 with 702 d.f.

3.5. The frequency distribution for the entire set of 10,400 deviates was obtained by pooling these 26 distributions, each for 400 entries. This distribution is given in Table 2. For purposes of application of the χ^2-test, the class-intervals were : $-\infty$ to -3.6, -3.6 to -3.4, …, 3.4 to 3.6, and 3.6 to $+\infty$. The value of χ^2 with 37 d.f. is 42.124, which indicates satisfactory fit.

TABLE 2. OBSERVED AND EXPECTED FREQUENCY DISTRIBUTIONS OF 10,400 NORMAL DEVIATES

class-interval	observed frequency (f_o)	expected frequency (f_e)
(1)	(2)	(3)
$-\infty$ to -4.0	0	1.655
-4.0 to -3.8	1	
-3.8 to -3.6	1	
-3.6 to -3.4	4	1.849
-3.4 to -3.2	4	3.642
-3.2 to -3.0	3	6.893
-3.0 to -2.8	15	12.534
-2.8 to -2.6	26	21.903
-2.6 to -2.4	40	36.778
-2.4 to -2.2	67	59.341
-2.2 to -2.0	93	92.006
-2.0 to -1.8	177	137.074
-1.8 to -1.6	211	196.238
-1.6 to -1.4	271	269.957
-1.4 to -1.2	336	356.855
-1.2 to -1.0	456	453.290
-1.0 to -0.8	537	553.281
-0.8 to -0.6	668	648.936
-0.6 to -0.4	775	731.382
-0.4 to -0.2	790	792.085
-0.2 to 0.0	842	824.301
0.0 to 0.2	826	824.301
0.2 to 0.4	755	792.085
0.4 to 0.6	713	731.382
0.6 to 0.8	590	648.936
0.8 to 1.0	547	553.281
1.0 to 1.2	480	453.290
1.2 to 1.4	349	356.855
1.4 to 1.6	272	269.957
1.6 to 1.8	185	196.238
1.8 to 2.0	129	137.074
2.0 to 2.2	97	92.006
2.2 to 2.4	69	59.341
2.4 to 2.6	35	36.778
2.6 to 2.8	19	21.903
2.8 to 3.0	9	12.534
3.0 to 3.2	3	6.893
3.2 to 3.4	3	3.642
3.4 to 3.6	1	1.849
3.6 to 3.8	1	1.655
3.8 to ∞	0	
total	10,400	10,400.000

χ^2 (37 d.f.) = 42.124

3.6. In addition to the above, it is possible to test the deviations of sample frequency constants from corresponding theoretical values. This has been done for the whole sample on the basis of the distribution presented in Table 2. Results shown in Table 3 seem to be fairly satisfactory.

TABLE 3. FREQUENCY CONSTANTS FOR THE ENTIRE SAMPLE (OF 10,400 NORMAL DEVIATES)

sl. no.	statistic		standard error	critical ratio
	symbol	value		
(1)	(2)	(3)	(4)	(5)
1	$\bar{x}$	-0.0223846	0.0098058	-2.283
2	s	1.0065689	0.0069337	0.947
3	β_1	0.0004949	0.0240192	0.021
4	β_2	2.9492832	0.0480384	-1.056
5	μ_3	-0.0226882	0.024019	-0.945
6	μ_4	3.0275446	0.096077	0.287
7	μ_5	-0.3091378	0.263117	-1.175
8	μ_6	14.6403318	0.988881	-0.364

4. TESTS BASED ON SAMPLE RANGES

4.1. The normal deviates, read vertically, were grouped into mutually exclusive pairs, and the absolute differences between them calculated. This gave the range for a random sample of size 2 from a standard normal population. Such calculations were done for the whole table, so that there were 200 samples per plate and 5,200 samples in all. The frequency distribution of each set of 200 ranges was compared with the theoretical distribution by the χ^2–test. Class-intervals were, after pooling: 0.0 to 0.2, 0.2 to 0.4, ..., 3.0 to 3.2, 3.2 to 3.6, 3.6 to ∞. The first line in Table 4 shows that 5 out of these 26 χ^2's were significant at the 10% level; only 1 crossed the 5% level but not the 1% level, and one value was beyond the 1% point. The total of these 26 values, which is itself a χ^2 with 442 d.f., was 450.365 only. The χ^2–test was also applied to the frequency distribution of all the 5,200 ranges taken together. First line of Table 5 shows that this was non-significant.

4.2. In exactly the same manner, frequency distributions of sample ranges were prepared for other sizes of sample, viz., n (size of sample) $= 3, 4, 5, 6, 7, 8, 9, 10, 11, 12$, and 15. (In each case, the process of forming samples started from the left-hand top corner of page 1, but as will be seen from cols. (1) to (3) of Table 4, for certain sample sizes the process did not cover the whole of the present 'Table'.) These observed distributions were compared with the corresponding normal theory distributions by the χ^2–test, which was applied (as in the case for $n = 2$) to the 'pooled' distributions of the entire set of sample ranges for any given n as well as to the distributions for suitable 'sub-sets' of 100 or 120 or 200 such sample ranges. As before, the χ^2's for the 'sub-sets' were combined by straightforward pooling. As a general rule, for $n = 3$ upwards, the 'minimum expectation' was taken to be 2 at the lower end and 1 at the upper. The results of all such tests are summarised in Tables 4 and 5. The over-all position seems to be satisfactory.

TABLE 4. RESULTS OF χ^2–TESTS ON FREQUENCY DISTRIBUTIONS OF 'SUB-SETS' OF SAMPLE RANGES

sample size (n)	no. of 'sub-sets'	no. of sample ranges per 'sub-set'	common d.f. for χ^2's	no. of χ^2's significant at different levels			pooled χ^2–test	
				10%	5%	1%	d.f.	χ^2–value
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
2	26	200	17	3	1	1	442	450.365
3	11	100	18		1		198	191.318
4	15	100	19	1			285	282.459
5	13	160	19		1		247	230.325
6	10	100	17	1		1	170	191.360
7	5	100	18	1			90	101.886
8	5	100	17	1	1		85	105.829
9	5	100	17		1		85	85.053
10	8	120	18				144	128.245
11	5	100	16				80	65.605
12	5	100	16				80	77.160
15	5	120	16				80	54.765

TABLE 5. RESULTS OF χ^2–TESTS ON 'POOLED' FREQUENCY DISTRIBUTIONS OF SAMPLE RANGES

sample size (n)	number of sample ranges in 'pooled' distribution	results of χ^2–test	
		d.f.	χ^2–value
(1)	(2)	(3)	(4)
2	5200	26	23.432
3	1100	24	26.740
4	1500	25	16.331
5	2080	27	23.285
6	1000	24	19.713
7	500	22	19.546
8	500	22	30.530
9	500	22	14.102
10	960	24	16.582
11	500	22	13.412
12	500	21	17.573
15	600	21	18.053

4.3. For each value of n, the mean of sample ranges ($\bar{R}_n$) was worked out from the 'pooled' frequency distribution of the entire set of ranges. These values were thereafter compared with the corresponding theoretical means (d_n), using exact values of standard deviation. Results shown in Table 6 seem to be fairly satisfactory.

TABLE 6. DEVIATION OF SAMPLE MEANS OF RANGES FROM THEORETICAL MEANS

sample size (n)	no. of sample ranges in 'pooled' distribution	observed mean range ($\bar{R}_n$)	theoretical mean (d_n)	critical ratio $(\bar{R}_n - d_n)/s.e.$
(1)	(2)	(3)	(4)	(5)
2	5200	1.1551	1.12838	2.260
3	1100	1.7313	1.69257	1.446
4	1500	2.0805	2.05875	0.957
5	2080	2.3557	2.32593	1.571
6	1000	2.5486	2.53441	0.529
7	500	2.7220	2.70436	0.473
8	500	2.8788	2.84720	0.862
9	500	3.0232	2.97003	1.472
10	960	3.0977	3.07751	0.785
11	500	3.1624	3.17287	−0.297
12	500	3.2768	3.25846	0.527
15	600	3.5170	3.47183	1.467

4.4. It should be appreciated that for the higher values of n, such tests are believed to be very sensitive to deviations from normality near the tails of the distribution and especially to the existence of what are called "rogue" observations in the sample.

5. TESTS OF RANDOMNESS

5.1. For the sake of interest, another series of tests based on the theory of runs was applied to the whole table. The theory of these tests will be found in Kendall (1946, Vol. 2, pp. 124-127). These tests, it must be noted, are tests of certain aspects of *randomness* and are not in any way concerned with *normality*.

5.2. The 104 sets of 100 deviates were considered separately. For each set, the signs of differences between successive values of normal deviates, when the table is read vertically, (one column after another,) were given. A sequence of n consecutive '+' or '—' signs, preceded and followed by the opposite sign, was taken as a phase of length n. The number of turning points i.e. peaks or troughs, was one greater than the number of phases.

5.3. If p be the number of turning points in a random series of N terms, the first four moments of p are known. For large values of N, the distribution of p is known to be approximately normal. For $N = 100$, in particular, we have $E(p) = 65.\dot{3}$, $\sigma(p) = 4.178$, $\beta_1(p) = 0.0005498$ and $\beta_2(p) = 2.9761$. Ignoring the slight departure from normality, the frequency distribution for the 104 values of p, one for each set, was prepared and compared with the normal distribution having mean $= 65.\dot{3}$, and s.d. $= 4.178$. The observed and expected distributions are shown in Table 7. Class-intervals were, after pooling : $-\infty$ to 56.5, 56.5 to 57.5, ..., 73.5 to 74.5, 74.5 to $+\infty$. The 'minimum allowable expectation' was taken to be 1. Goodness of fit χ^2 came out to be 25.639, which for 19 degrees of freedom is non-significant.

TABLE 7. OBSERVED AND EXPECTED[1] FREQUENCY DISTRIBUTIONS OF p, THE NUMBER OF TURNING POINTS IN SETS OF 100 DEVIATES

no. of turning points (p) (1)	observed frequency (f_o) (2)	expected frequency (f_e) (3)
upto 51	0	0.049
52	1	0.063
53	0	0.131
54	2	0.257
55	0	0.477
56	0	0.836
57	0	1.385
58	2	2.055
59	5	3.146
60	5	4.396
61	5	5.799
62	11	7.224
63	13	8.499
64	5	9.440
65	6	9.902
66	6	9.808
67	7	9.174
68	7	8.102
69	14	6.757
70	5	5.321
71	2	3.957
72	5	2.779
73	2	1.781
74	1	1.178
75 & above	0	1.483
	104	104.000

---

[1] Expected frequencies were calculated on the assumption of normality, with $E(p) = 65.\dot{3}$, and $\sigma(p) = 4.17798$.

5.4. For each set of 100 deviates, the frequency distribution of observed phase lengths was prepared. Phases of length 3 or more were pooled together. The use of such frequency distributions for tests of randomness has been studied by Wallis and Moore (1941). The probability of a phase of length d in a series of length N is given (approximately) by the formula $\frac{6(d^2+3d+1)(N-d-3)}{(d+3)!(2N-7)}$.

The expected frequencies were adjusted so as to add up to the observed number of phases. If χ^2 is then calculated in the usual manner, one has to consider 2.5 as the appropriate d.f. if $\chi^2 \geqslant 6.3$. For lower values of χ^2, $\frac{6}{7}\chi^2$ can be tested as a χ^2 with d.f. 2.

5.5. Such tests were carried out for all the 104 distributions. In 7 cases the values were significant at the 10% level; out of these 7 values, 2 exceeded the 5% level, but none exceeded the 1% level.

6. Acknowledgement

6.0. The authors are glad to acknowledge the help received from R. K. Som, P. Chowdhury, K. C. Poddar, S. N. Haldar and S. K. Roy in the matter of the revision of the tables; from T. R. Talwar, A. K. Bhattacharya and R. N. Mukherjee in connection with the statistical tests carried out on the tables; and from I. M. Chakravarti, who made valuable suggestions in the preparation of the introductory note.

References

Cochran, W. G. (1952) : The χ^2 test of goodness of fit. *Ann. Math. Stat.*, **23**, 315.

Das, A. C. (1951) : On the estimation of parameters in a recursive system. *Sankhyā*, **11**, 273.

Hojo, Tokishige (1931) : Distribution of the median, quartiles and inter-quartile distance in samples from a normal population. *Biometrika*, **23**, 315.

Kelley, T. L. (1948) : *The Kelley Statistical Tables*, Revised 1948, Harvard University Press.

Kendall, M. G. (1946) : *The Advanced Theory of Statistics*. Charles Griffin & Co. Ltd., London.

Kendall, M. G. and Smith, B. Babington (1939) : *Tables of Random Sampling Numbers*. *Tracts for Computers*, No. XXIV, Cambridge University Press.

Mahalanobis, P. C., Bose, Subhendu Sekhar, Roy, Prabhat Ranjan and Banerjee, Sudhir Kumar (1934) : Tables of random samples from a normal population. *Sankhyā*, **1**, 289.

Matthai, A. and Kanan, M. B. (1951) : The applicability of large sample tests for moving average and auto-regressive schemes to series of short length—an experimental study. Part I : Moving average. *Sankhyā*, **11**, 218.

PEARSON, E. S. (1932) : The percentage limits for the distribution of range in samples from a normal population. *Biometrika*, **24**, 404.

PEARSON, E. S. AND HARTLEY, H. O. (1942) : The probability integral of the range in samples of n observations from a normal population. *Biometrika*, **32**, 301.

PEARSON, E. S. AND WISHART, JOHN (1942) : *"Student's" Collected Papers* (Edited).

PEPPER, JOSEPH (1932) : The sampling distribution of the third moment coefficient—an experiment. *Biometrika*, **24**, 55.

RAO, S. R. AND SOM, R. K. (1951) : The applicability of large sample tests for moving average and auto-regressive schemes to series of short length—an experimental study. Part II : Auto-regressive series, *Sankhyā*, **11**, 239.

RIDER, PAUL R. (1932) : On the distribution of the correlation coefficient in small samples. *Biometrika*, **24**, 382.

THE RAND CORPORATION (1955) : *A Million Random Digits with 100,000 Normal Deviates*, The Free Press, Publishers, Glencore, Illinois.

TIPPETT, L. H. C. (1927) : *Random Sampling Numbers. Tracts for Computers*, No. XV, Cambridge University Press.

WALLIS, W. A. AND MOORE, G. H. (1941) : A significance test for time series. Technical paper No. 1, *Nat. Bur. Ec. Research.*

WOLD, H. O. A. (1948) : *Random Normal Deviates. Tracts for Computers*, No. XXV, Cambridge University Press.

Paper received : April, 1958.

TABLES OF RANDOM NORMAL DEVIATES

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 1

col. no. row no.	1	2	3	4	5	6	7	8
	2 0538	1 0424	2 0255	1 2037	1 0834	1 0230	2 0476	1 0157
	2 0210	1 1669	2 1017	2 1082	1 0584	1 0089	2 1126	2 0615
	2 0604	1 0669	2 0410	2 0594	2 0368	2 1232	1 0500	1 0736
	2 1589	1 0062	2 1220	1 0281	1 0255	1 0888	2 0193	1 1163
5	2 0597	1 1368	2 1078	1 1296	1 0525	1 0282	1 1183	1 0371
	1 0220	2 0565	2 0003	2 0971	2 0547	1 0297	1 0271	2 0594
	1 1449	2 0685	2 0413	1 0111	1 1145	1 0261	1 0012	2 0618
	2 0841	1 0786	2 0958	1 0676	2 1725	1 1007	1 1155	1 0556
	1 1715	2 0572	1 1578	2 0340	2 0645	1 1185	2 0858	1 0389
10	1 0932	2 1003	2 0867	1 0007	2 0405	1 0550	2 0256	1 0568
	1 0168	2 0382	1 1454	1 0331	1 0357	1 0615	1 0806	1 0786
	2 1271	1 0493	2 1169	1 0402	2 0772	1 0003	2 1454	2 0705
	1 0479	1 0028	1 0173	2 0312	1 0629	2 0403	1 0968	1 0365
	2 0827	2 0770	2 0028	1 0627	2 0564	2 1015	1 0243	1 0120
15	1 0142	2 0059	2 0638	1 0070	2 0889	1 0385	1 0188	1 1723
	2 0124	2 0912	2 0443	2 0255	1 1630	2 0192	2 0572	1 2612
	2 0789	1 0378	1 0174	1 1483	1 0726	2 0967	1 0108	1 0725
	1 0501	2 0224	1 0625	1 0483	2 1189	1 0591	2 1136	1 0021
	1 0931	2 0221	2 0220	1 1645	1 0165	2 0271	1 0238	2 0435
20	1 0811	2 0329	1 0239	1 1983	2 0596	2 0199	2 0001	1 1106
	1 1542	1 0561	2 0459	2 1449	2 0537	2 1274	1 0350	2 0127
	2 0403	1 0330	1 0260	1 1541	2 0428	2 1242	2 1050	2 0050
	1 1337	2 0981	1 0018	2 1503	1 0388	2 1330	2 0100	2 0178
	1 1423	1 1473	2 0584	1 0553	2 0239	2 0815	1 0331	2 0647
25	1 0766	2 0316	2 0555	1 0723	2 2362	1 0528	2 1123	2 0861
	1 1946	1 1872	1 0627	2 2929	1 1719	2 0896	2 2270	2 0878
	1 1572	1 1412	1 1169	1 1534	2 0085	2 1755	1 0445	1 0141
	2 1186	2 0366	1 0251	2 0508	1 1290	1 0153	2 0723	1 0894
	2 1470	2 0251	2 0239	2 1015	2 0964	2 1091	1 0061	2 0144
30	1 0345	2 0254	2 0307	2 0780	1 0909	2 0122	1 0345	2 0390
	2 1074	1 0569	2 0343	2 0980	2 1524	2 0401	2 0141	2 0500
	1 0536	1 1273	1 0528	1 0170	1 0697	1 0436	2 0925	2 0481
	2 0456	2 0588	2 1379	1 1313	1 1207	1 0043	1 1008	1 1351
	1 0814	2 0017	1 0190	1 0295	1 0403	1 1081	2 0406	1 1325
35	1 1160	2 0382	1 1268	2 1419	1 0354	2 2759	2 0637	1 0249
	2 0671	1 0605	1 1147	2 3891	2 0896	2 0704	2 0358	1 0045
	2 2007	2 0705	1 2077	1 0424	1 1930	2 0969	1 1377	2 1079
	2 0522	1 1043	1 0596	1 1562	2 0570	2 1463	2 1242	2 0485
	1 0671	2 2012	1 0814	2 0722	2 0614	2 0024	2 0680	2 0288
40	1 0047	2 0906	2 0765	1 1540	2 1615	2 0872	2 0919	2 0813
	2 1176	2 0660	2 0680	1 1759	2 2473	2 0320	1 2214	1 0024
	2 0277	2 0081	1 0778	1 0390	1 1354	2 0476	2 0223	2 0354
	2 1230	2 0764	2 0964	2 0745	2 0329	1 0912	2 1112	2 1056
	2 0909	1 0746	1 0146	1 0569	1 1663	2 1994	1 0766	1 0187
45	1 0308	1 1748	1 1778	2 0802	1 0754	2 0808	1 0007	2 1588
	2 0062	2 0121	2 0598	1 0849	2 0090	2 1180	1 0613	1 0974
	2 1659	1 0522	2 0492	1 0006	2 0028	2 1006	1 2095	1 0469
	2 0015	2 0523	2 1062	2 0580	2 1045	2 1299	1 0613	2 0312
	2 0018	1 2024	1 0662	2 0254	2 0564	2 1590	2 0257	1 0875
50	2 0723	2 0997	1 0569	2 1948	2 0243	2 0226	2 0473	1 2066

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 2

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 1148	2 0566	1 0630	2 0247	2 0336	1 1030	2 0638	1 0123
	1 0111	2 0022	2 1327	1 0563	2 0051	2 1121	1 1127	2 0769
	2 0450	2 0732	2 0224	1 2248	2 0056	2 0123	2 0936	2 0646
	1 0775	2 1885	1 1257	2 0786	1 0765	2 1437	2 1159	1 0577
5	1 0485	2 0203	2 0360	1 0018	2 0132	2 0496	2 1455	2 0013
	1 1391	2 0302	1 1374	2 0797	1 0452	1 1807	1 0880	2 0487
	1 0080	2 1290	2 1528	1 0380	2 0042	1 1131	1 1190	1 1068
	2 2238	2 0056	1 0854	1 1417	2 0797	2 0761	1 0010	1 1097
	2 0715	1 0763	1 1595	1 0576	1 0996	2 0261	2 1341	1 0069
10	2 1803	1 0069	2 0254	2 1941	1 1443	1 0197	1 0184	2 2727
	1 0791	1 1891	1 0065	2 0186	1 2251	2 0021	2 1289	1 1352
	2 0671	2 0188	2 1018	2 2007	2 2284	1 0392	1 1111	2 0074
	1 0423	2 0873	2 0490	1 1372	1 0291	1 0226	2 1457	2 0284
	1 0459	2 1065	1 1345	2 0060	1 2334	2 0315	1 0640	2 0604
15	2 0119	1 0842	1 1102	2 0422	1 0822	2 0038	2 0172	1 1794
	1 0176	1 2267	2 0688	2 1345	2 0219	1 0431	2 1406	1 0289
	2 0212	2 1060	2 0557	2 1458	2 1129	1 1275	2 1025	2 1051
	2 0124	2 0683	2 0361	1 0077	1 0719	1 0562	2 1269	2 0701
	2 0089	2 1544	1 2160	2 0136	1 0168	1 0729	2 1183	2 0730
20	1 0436	2 0373	1 2115	1 1396	2 1639	1 0378	1 0089	1 1248
	1 0386	2 1840	1 0806	1 0920	2 1371	1 0865	1 0178	2 0863
	2 0971	2 0179	2 0894	1 0080	2 0382	1 1271	1 0664	2 0296
	2 1117	2 1173	2 0428	1 1219	2 0267	1 0461	2 0954	2 1848
	1 0008	2 0246	2 0814	1 2079	1 1101	2 0443	1 0580	1 0293
25	1 0795	2 0831	1 0039	2 0841	2 0542	2 1983	1 0729	2 0567
	2 0848	1 1277	2 1272	1 0602	1 0739	2 0214	1 0890	1 0604
	1 2536	2 1375	1 0664	2 0270	1 0842	2 1368	1 0198	2 0131
	2 1838	1 2378	2 0003	1 0769	1 0505	2 1359	2 0846	1 2268
	1 2338	1 0811	2 0218	1 0475	2 1995	2 1109	1 0647	1 1190
30	2 0989	1 0793	1 1645	2 0225	2 1261	2 0531	1 0273	2 1295
	2 0808	2 1193	2 0309	1 0320	2 1528	2 1481	2 0751	1 0073
	2 1634	2 1997	1 0241	1 0470	2 1076	1 0714	2 0024	2 0077
	2 1622	2 0239	1 1366	1 0351	2 0807	1 0950	2 0356	1 0386
	1 0637	1 0639	1 0404	2 1008	2 0212	2 1718	2 0389	2 0321
35	2 0437	2 0323	1 0800	1 0520	1 0797	1 0602	2 0104	1 1740
	1 0488	2 0664	1 0670	1 1746	1 1332	2 0613	2 0008	2 0999
	1 1274	2 0490	2 0351	2 1784	2 1249	1 0342	1 1266	2 0197
	1 1408	2 0720	2 1759	2 0469	2 0990	1 0993	2 0154	2 2081
	1 1958	1 0836	1 0546	1 2036	2 1024	1 0888	2 0132	2 0880
40	2 0778	2 0803	2 0561	1 0185	2 1008	1 0760	2 0015	1 2404
	1 0503	2 0500	2 0814	2 0432	2 0143	1 1669	2 1843	1 1799
	2 0420	2 0367	2 1785	1 0236	2 0273	1 0598	1 1070	1 0220
	1 1997	1 0136	2 0071	1 0132	2 0315	2 0477	2 0337	2 0151
	1 1516	1 0145	1 0119	2 2075	1 0605	1 0522	2 0274	1 0604
45	1 0192	1 0332	2 0370	1 0307	2 0521	1 0304	1 1020	2 0825
	1 0482	1 0886	2 0137	2 0542	1 2543	1 1863	2 0362	1 0959
	1 0417	1 0457	2 0016	2 2478	2 1761	1 1225	2 0062	1 0644
	2 0633	2 0188	1 1153	1 1009	2 0168	1 1277	2 1146	2 1601
	1 0478	2 0026	1 1672	2 0378	1 0395	1 0740	1 2518	1 1017
50	2 0120	1 0763	2 0282	1 0596	2 1112	1 0558	2 0026	1 0417

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE NO. 3

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0996	2 0505	1 0809	1 0428	1 0075	1 0023	1 1071	1 0455
	2 0012	1 0522	1 0061	1 0352	1 0022	2 0814	1 0970	2 1849
	1 2217	2 1752	1 0451	2 0650	1 0691	2 0209	2 1097	1 1546
	2 0622	1 0831	2 0551	1 1129	2 2079	1 2071	1 0585	1 0879
5	1 0204	2 0169	1 0435	1 2175	1 1605	2 1364	1 1033	2 1192
	1 0953	2 1213	1 1374	1 1662	2 1045	2 1354	2 0721	2 0639
	2 0491	2 1979	2 1426	1 0854	1 2848	1 1353	2 1301	2 1833
	2 0711	2 0849	2 1216	2 0493	2 1191	2 2167	2 0287	2 0423
	2 2661	2 0136	1 1114	1 0074	1 0882	1 0189	1 0659	2 0792
10	2 0706	2 0007	2 1681	2 0897	1 0540	2 0498	1 0485	1 0764
	1 0251	1 0453	1 1191	2 0595	1 0055	2 0428	1 0177	1 0348
	2 1284	2 1372	1 1824	1 0555	2 2254	1 0056	2 1163	2 0030
	1 0498	2 0007	2 0231	2 0054	1 1872	1 1138	2 0231	2 0247
	1 0786	1 0235	1 0899	2 2226	1 1184	1 0172	1 0058	1 0446
15	2 0027	1 0312	1 1441	1 1446	2 0781	2 1875	2 1373	1 0366
	2 1261	1 1361	1 0102	2 0941	1 0029	2 0269	2 1252	1 0308
	1 1366	1 0805	2 1353	1 0543	1 0218	2 0685	1 1985	1 0811
	2 1135	1 1086	2 0404	2 2312	2 1484	2 0321	2 0053	1 2280
	1 1577	2 0300	1 0192	1 0848	2 0349	1 1417	2 1150	2 0896
20	1 1376	2 0033	1 1453	1 0292	2 0133	1 0844	1 0039	2 0801
	1 0710	2 0383	1 1393	2 0939	1 0195	1 0009	1 1441	2 1332
	1 0011	2 0379	1 0935	2 0240	2 0165	2 0462	1 1408	2 0056
	1 0076	2 0625	2 0987	2 0997	2 1820	1 0467	2 0283	2 1580
	2 0335	1 0007	1 0332	2 0113	1 1658	1 0219	1 0018	2 1071
25	2 2175	2 0691	1 0123	2 0641	1 0107	2 1866	1 1985	2 0887
	2 0559	1 0784	2 1181	2 0952	2 0754	2 0167	2 0033	2 1504
	1 0815	1 0127	2 1715	2 0496	2 1284	2 0600	2 1007	2 0672
	2 0849	1 1003	1 0590	1 0951	2 0657	1 0949	1 2834	2 0023
	2 1289	2 0685	2 0908	2 0333	1 0722	2 0533	2 0848	2 0492
30	1 0693	1 0142	2 0696	2 2186	2 2330	1 1726	1 1477	1 1238
	1 1372	2 1237	2 1721	2 0202	2 0847	2 1243	2 0498	2 0792
	1 0476	1 0486	2 2727	1 2008	2 0995	1 1171	1 0226	2 0266
	2 0335	2 0092	1 1200	1 0600	2 1027	2 0325	1 0975	1 0827
	1 0335	2 2442	2 0611	2 1591	2 0904	2 0327	1 0163	2 0724
35	2 0126	2 0980	1 1815	2 0292	2 0523	2 1199	1 1885	1 0225
	2 0632	2 0702	2 1720	2 1087	1 0535	1 1311	2 1721	1 0133
	1 0222	2 1228	1 0924	2 0666	2 0027	1 0283	1 0247	1 1009
	2 0195	2 0523	1 0530	2 1099	1 2073	2 0219	2 0154	1 0860
	2 0877	2 1585	1 1631	2 0484	2 1773	2 1239	2 0174	1 0845
40	1 0259	2 1412	2 0311	1 1165	2 0147	2 1238	2 0347	1 0197
	1 0503	1 0590	1 0537	2 1274	1 1017	2 0745	2 0145	2 0600
	1 0310	2 1536	1 0813	2 1079	1 1529	2 1095	2 0376	1 1537
	2 0490	1 0419	1 0947	2 1194	1 0177	2 0402	2 0542	2 1081
	2 0297	1 0037	2 1802	1 0601	1 1795	2 1696	1 0899	2 0593
45	1 2304	2 0085	2 0711	2 1201	1 0182	2 0275	1 1074	1 0389
	1 1074	1 0007	2 0172	2 1548	2 1208	1 0047	1 1318	2 0044
	2 0850	1 0814	2 0278	1 1179	2 0797	2 2290	2 0136	2 0552
	1 0165	2 0450	2 0303	1 0962	2 0660	1 1900	2 1188	1 1367
	1 0192	2 1421	2 0686	2 0921	2 0293	1 1545	2 0173	0000
50	1 1798	2 2669	1 0553	1 1167	1 1307	2 1779	2 0117	1 0470

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE NO. 4

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0841	2 0262	1 2120	1 0165	2 2304	2 0274	1 0961	1 2473
	1 0140	2 1082	2 0988	2 0407	2 0474	2 2370	1 1070	2 0304
	2 1437	2 0496	1 1256	1 0004	2 0674	2 0336	2 0675	1 0379
	2 1015	2 1722	2 0612	1 0254	1 1159	2 1612	2 0580	1 0172
5	1 2280	2 0380	1 0748	2 1109	1 2304	2 0129	2 0733	1 0055
	2 0225	1 0204	2 0818	2 0140	2 1596	2 1125	1 0097	2 0568
	1 1872	1 0160	1 0586	1 0344	2 1128	2 0057	2 0188	2 0396
	2 0502	1 0705	1 0654	2 1231	1 0512	1 0354	1 0421	1 0295
	2 0583	2 0481	2 1547	1 0118	2 0009	2 0050	1 1133	1 1522
10	2 1720	1 1551	1 1643	1 0947	1 0917	2 0101	1 0352	1 1363
	1 0189	2 0449	1 0247	1 2042	1 0440	1 0403	2 0078	1 0092
	2 1730	1 1063	1 1200	1 0991	1 2088	1 0626	1 1567	1 1323
	1 1043	2 0075	2 0726	1 0416	1 0912	2 0452	1 0183	1 0813
	1 0773	2 0165	1 0678	1 1932	2 0401	1 0907	2 0283	2 0639
15	2 0197	2 2330	2 0013	2 1808	2 0202	1 0883	1 1275	1 1454
	2 1102	2 0534	1 2125	1 0905	1 0176	1 0631	2 0670	2 0806
	2 1885	1 1152	1 0675	1 0464	1 0440	2 0391	2 0016	2 0088
	2 1146	1 0143	2 0219	1 0529	1 0274	2 0072	1 0315	2 0345
	2 0248	1 2052	2 1048	2 1811	1 0226	2 0073	1 1338	1 0024
20	1 0984	2 0253	2 0709	2 1102	1 1782	1 1379	2 0479	2 0034
	1 1068	1 1309	1 0325	1 0093	2 0957	1 1099	2 0069	1 0357
	1 0230	1 0455	2 0534	2 1652	1 0514	2 1321	2 1616	1 0312
	2 0731	2 1322	2 1118	1 1037	2 1759	1 0629	1 0239	1 1851
	2 0634	2 1323	1 1958	1 1678	2 0661	2 1570	2 0605	2 0191
25	2 0463	1 0387	2 0721	1 0237	2 0489	2 0542	1 0403	1 0739
	2 1298	1 0601	1 0268	2 1604	2 0121	2 1044	1 0394	1 0681
	1 2326	2 0288	1 1513	2 1623	1 0439	1 1920	2 0229	2 0882
	1 0411	2 0861	2 0903	1 0523	2 0624	2 0444	1 0141	2 0694
	2 0794	2 0275	1 0442	2 0625	1 0694	2 0817	1 1363	1 0887
30	1 0606	2 0374	2 0890	1 0299	1 0152	2 0446	2 0027	2 1342
	2 0802	1 1554	2 0758	2 0792	1 1138	2 1169	1 1630	2 0257
	1 0607	2 2226	1 0769	1 0324	2 1060	1 1069	1 0494	1 0030
	2 0510	1 0240	1 0569	2 0600	2 1259	2 1974	1 1117	2 0175
	2 0635	1 0752	1 0438	1 0252	1 0680	2 1108	1 0471	2 0218
35	1 0507	1 0877	1 0290	2 0837	1 1187	2 1104	2 0814	2 2432
	1 1574	1 1480	2 0092	1 0715	1 0190	1 0325	2 0899	2 0276
	2 0025	1 0229	2 0467	2 0080	1 1266	2 0781	1 0456	2 1077
	1 0450	2 0481	1 0913	2 0412	1 0179	2 0334	1 2073	2 0467
	2 0831	2 2152	1 1249	1 1356	1 1370	2 0728	1 0210	2 1242
40	1 0033	1 0842	2 0507	1 0096	2 0698	1 0179	1 0902	1 0498
	1 0304	1 0759	2 0240	2 0726	1 1505	2 0111	1 0194	2 2549
	1 0797	2 2165	2 2506	1 0954	2 0429	1 0901	2 1399	1 0459
	1 0981	1 0582	2 0819	2 0136	2 0669	2 0888	1 0041	2 0778
	2 0133	1 0157	1 0489	2 0984	1 1195	1 0067	1 0048	2 0431
45	2 0378	1 0376	1 0488	2 2024	1 1328	2 0055	1 1220	1 0157
	1 0946	2 1192	2 0475	1 1359	1 0288	2 0964	2 0503	2 0339
	1 0834	2 0123	1 0058	1 0889	1 1019	2 2200	2 1305	1 0748
	1 0532	2 0005	2 1100	2 1128	1 0832	1 0364	2 0809	2 0463
	2 0307	2 2484	1 0368	1 0075	1 0068	2 0177	2 0745	1 1244
50	1 1011	1 1320	2 1276	2 0565	1 1569	2 1279	2 0522	2 0953

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

TABLES OF RANDOM NORMAL DEVIATES

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 5

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0894	2 0491	1 0282	2 2066	2 0403	2 1826	1 0434	2 0097
	1 0267	1 0211	1 0975	1 0402	2 0104	2 0086	1 1476	2 0017
	2 1551	2 1354	2 0635	2 0516	2 0148	2 0212	2 0155	2 0411
	2 0308	1 0434	2 0343	2 0507	2 0590	1 0795	2 0098	2 1137
5	1 0265	1 0181	2 0618	1 0201	1 0859	1 0298	1 0216	1 0868
	2 0760	1 2099	1 0356	1 0264	1 1390	1 0390	1 0570	1 0523
	2 0617	2 0599	2 2122	1 1265	1 0375	2 0402	2 0396	2 0724
	2 2077	1 0566	1 0046	2 1269	1 1243	2 0010	1 1237	2 1229
	1 0748	1 0863	2 0242	2 0481	1 0962	2 0645	2 1542	1 1134
10	2 0666	1 1582	2 1005	1 0252	1 2911	2 1435	1 0994	1 0078
	2 2020	2 1875	1 0296	1 0040	2 0961	2 0314	1 0289	2 1199
	2 1294	1 0453	2 0688	1 1304	2 1672	2 1680	2 0468	2 1006
	2 0674	1 0606	2 1644	0000	2 0627	1 0426	1 0885	2 0194
	1 2113	1 0503	1 0634	2 1142	1 0955	1 1396	2 0757	1 0314
15	1 0087	1 0331	1 1427	2 0979	1 0849	1 0781	2 0809	1 0273
	1 0161	1 0204	2 1009	2 0155	2 0841	1 0274	1 0380	1 0501
	1 1621	2 0159	1 1690	1 0469	1 1630	1 1014	1 1315	1 1421
	1 1580	1 0279	2 0900	1 0795	2 0072	2 1172	1 2235	1 0537
	2 1344	1 0594	1 0094	1 0401	1 1284	1 1932	2 0006	1 2428
20	1 1077	1 0185	1 0087	1 0143	2 0228	2 1644	1 0556	1 0433
	1 1852	2 2245	1 0457	1 0387	2 0630	2 0474	2 0810	1 0206
	1 0988	2 1213	1 0056	2 0711	2 1312	2 1166	1 0757	1 1583
	2 1776	2 0447	2 0566	2 0409	1 0383	1 0645	1 0737	2 1899
	1 0268	2 1366	2 0145	1 0260	2 1083	2 0194	1 1299	2 1036
25	1 0009	1 0942	2 0312	1 0102	1 1358	2 0489	2 0504	1 1630
	2 0609	2 0910	2 1596	2 1744	2 1457	2 1193	1 0379	2 0616
	1 1030	1 0166	2 0430	2 0296	1 0508	2 1084	2 0139	1 1364
	2 0821	1 0404	1 0086	1 1079	2 0195	2 0337	1 1832	2 0363
	2 0584	1 0631	2 2056	1 0902	1 0907	1 0056	2 0287	2 0078
30	2 0200	2 0557	2 0043	2 0767	2 1140	2 1624	1 0686	2 0250
	2 1385	1 0454	1 0049	2 0732	1 1604	2 1165	1 0464	2 0406
	1 0152	1 1751	2 0474	2 0951	2 0275	1 1396	1 2022	2 1570
	2 0483	1 1026	1 0450	1 2092	1 0937	2 0431	2 0746	1 0447
	2 0487	1 0154	1 0087	2 0707	2 0352	2 1630	2 1031	1 0864
35	2 0148	1 0299	2 0052	2 2848	2 0018	2 0856	2 0404	2 0469
	1 1005	1 0483	1 2127	1 0451	2 1193	1 0006	2 1131	2 0443
	1 0716	2 1738	2 0351	2 0542	1 0362	2 0853	1 0044	2 0597
	2 0776	2 1270	1 0109	2 1568	2 0623	1 1254	2 1063	2 0445
	1 0375	2 0131	1 0203	1 1534	2 0951	2 0666	2 0010	1 1217
40	2 0544	1 1499	1 0019	1 2727	2 1306	2 1380	1 0412	2 1266
	2 2173	2 0449	1 0544	2 0196	2 1274	2 2092	2 0411	2 0035
	1 0923	1 1251	2 1639	2 0131	2 0166	1 1074	2 0056	1 0270
	2 0220	1 0279	1 0881	1 0613	1 0584	2 0180	2 1108	1 1903
	2 0331	1 0038	2 0852	1 0385	2 0460	1 1688	1 1004	2 0868
45	2 0892	1 1449	2 2748	1 0811	1 1194	1 0012	2 0677	2 0407
	2 1239	1 0577	1 0578	1 1261	1 0821	1 0564	1 0063	1 0870
	1 1227	2 0110	2 0893	2 1121	1 0078	1 0603	1 0871	1 0274
	1 0196	2 0491	1 0655	2 0307	2 1581	1 1345	2 0562	1 0065
	1 0547	2 0577	1 0840	1 1370	1 0800	2 1306	2 0069	2 1718
50	1 0125	2 0078	1 0137	2 0120	1 0682	2 1920	2 0298	2 0127

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 6

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 1094	2 0159	2 0284	1 1146	2 1165	2 0224	1 0738	1 1048
	2 0613	1 0946	1 2346	2 0739	1 0058	1 0875	2 0292	2 1330
	1 0599	1 1729	2 1406	2 1048	1 0065	1 0616	2 0230	1 1311
	2 0769	2 1361	2 0328	2 0342	2 0841	2 0359	2 0138	2 0710
5	2 0785	1 1312	1 1068	1 1124	1 1135	1 0003	2 0489	2 0762
	2 0066	1 0142	1 0024	1 1849	2 0400	1 0070	2 0403	2 0809
	1 1836	2 1123	2 0488	1 0527	2 0968	1 0881	2 0355	2 0953
	2 0102	1 1509	1 1161	2 0620	1 0927	1 0603	2 0659	2 0234
	1 0801	1 0567	2 0463	1 0200	1 0051	1 1491	1 0594	2 0413
10	1 0050	1 0022	1 0320	1 1311	2 2200	2 0889	2 0946	1 0675
	1 0457	2 0524	1 1184	2 0325	2 0180	2 2727	1 0693	2 0312
	2 1869	1 1269	2 1155	2 1026	1 0058	1 1484	2 1104	2 0375
	2 0123	2 1679	1 1616	2 0420	1 1916	2 0228	1 0633	1 0142
	1 0318	2 0076	2 1710	2 0488	2 0529	1 1052	2 0994	2 0549
15	1 0243	2 0423	1 0086	1 0261	2 1002	1 0456	2 0026	2 1113
	1 0602	1 0445	2 0689	1 0477	2 0681	2 1084	1 0633	2 0200
	2 0093	2 1007	2 0554	1 0084	2 1764	1 0984	2 0740	2 1521
	2 0805	2 1331	2 0482	2 1635	1 2203	2 1074	1 0183	2 1212
	1 1029	1 0085	1 1415	2 1416	2 1750	1 0417	2 0056	1 0277
20	1 0696	2 0230	2 0255	2 0098	1 1615	2 0277	2 0473	2 0500
	2 0182	2 0530	1 0624	1 1413	1 0905	1 0022	1 1375	2 0041
	1 0346	2 1291	1 0428	1 1414	1 1902	1 1139	1 0982	2 1231
	1 0108	1 0528	1 1078	1 0832	1 0755	1 0200	2 0572	1 0080
	2 0177	1 0015	2 0011	2 0291	1 1061	1 1040	1 1122	1 0330
25	1 0058	2 0057	2 0627	2 0913	1 0704	1 1653	2 0679	1 1104
	1 1088	1 2203	2 0135	1 0445	2 2024	1 0819	1 1294	1 0260
	2 0408	2 1272	2 0159	2 0073	1 0414	1 1875	2 0879	2 1953
	2 0712	2 0668	2 0489	2 1321	1 0190	1 1177	1 1730	1 0176
	1 0370	2 1442	2 0857	2 0598	2 0743	2 0204	2 0055	1 0673
30	2 0583	1 0063	2 0678	1 0484	1 0543	1 0617	2 0659	1 0530
	1 0621	2 0018	1 0581	1 0106	1 1098	1 0079	2 0437	1 0836
	1 0495	1 0920	1 0542	1 0177	2 0596	1 0585	1 1166	2 0063
	2 0018	2 0520	1 1436	1 1845	2 1621	2 1228	1 0060	1 1039
	1 0912	1 1100	2 0931	1 0725	1 1797	1 1812	2 0369	1 1329
35	1 0528	1 1327	1 0108	1 0679	2 0293	1 1389	2 1063	2 2652
	1 0726	2 0371	2 0056	2 0007	1 0252	2 1080	1 2226	1 0426
	2 0467	2 0239	2 1579	2 0987	1 0120	1 0280	1 2122	2 1372
	2 0582	2 0441	2 0461	2 0062	2 0790	2 0119	2 0917	2 1010
	1 1137	2 0390	2 2044	1 0388	1 0442	1 0324	2 0435	1 0435
40	2 0978	1 0221	1 0723	2 0318	2 0754	2 0205	1 0156	1 0056
	2 2280	2 0504	2 1833	2 0522	2 0534	1 0202	1 0411	2 2770
	2 1679	2 0023	1 0872	1 0082	1 0747	1 0694	2 0871	1 1186
	1 0351	2 0724	1 1403	1 0102	1 1183	2 0115	2 1218	1 1126
	2 1571	1 0899	2 1652	1 1400	1 1580	2 0010	2 1093	2 1051
45	2 0885	1 0659	1 0097	1 0098	1 1125	1 0141	1 0202	1 0689
	2 0751	2 0098	2 0771	2 0306	1 0765	1 0982	2 0475	1 1314
	1 0062	2 0491	1 0507	1 0737	2 0783	1 1296	1 1799	1 0019
	1 0224	2 1527	1 0519	2 0890	1 0393	1 1327	2 0535	1 0968
	1 1128	2 1687	2 2418	2 1321	1 0181	2 0175	2 0101	2 0805
50	1 1079	1 0292	2 0642	2 0169	2 1189	1 0627	2 1386	2 0216

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 7*

row no. \ col. no. 1		2	3	4	5	6	7	8
	1 0886	1 0687	2 0157	1 0871	1 0599	2 0431	1 0110	2 1201
	2 0206	2 1117	1 0429	2 0765	1 1487	2 1225	2 1041	2 0815
	1 0650	1 0123	1 0999	2 0963	2 2077	1 0972	2 0723	1 0797
	1 1374	1 1125	2 0181	1 0969	2 0469	2 1170	1 1023	2 0531
5	2 0987	2 0458	2 3195	2 0658	1 0833	1 0762	1 0719	2 0968
	1 0882	2 1024	2 0177	1 0454	2 0298	2 1134	1 2220	1 0092
	1 0992	2 0034	1 0008	1 0200	1 0382	1 0143	1 0585	2 0837
	1 1043	1 1084	2 0751	2 1524	2 0964	1 0561	1 0520	1 0620
	2 0343	1 1335	2 0709	2 0876	2 0982	2 0645	1 0394	2 0514
10	1 1059	2 1521	2 0531	1 0029	2 0025	2 0616	2 0106	1 0010
	1 0457	1 0537	1 0174	2 1854	1 0418	1 0514	2 0604	1 0817
	2 1029	1 1779	1 1653	2 0265	2 1022	2 0860	1 2111	2 1106
	2 0979	1 0221	2 0533	2 0010	2 1113	1 1329	1 1527	2 0138
	2 0963	2 0139	2 0411	2 0076	1 1213	1 0026	2 0263	1 0633
15	2 0852	1 2229	2 0859	1 1118	1 1743	1 1000	2 1940	2 0280
	2 0255	1 0168	2 0344	1 0137	2 1099	1 0979	1 0431	2 0966
	1 0145	2 1365	1 0130	2 1005	1 0763	1 0386	2 0182	1 1517
	1 2088	1 1053	1 0453	2 1995	2 0061	1 0443	1 0236	2 0397
	2 0746	1 0543	1 0777	1 1946	1 0898	2 1250	2 0198	1 1194
20	2 1370	2 0256	2 0293	1 0650	2 1092	2 0336	1 0055	2 0624
	2 0631	2 0927	2 0255	2 1575	2 1127	1 0412	1 0649	1 1202
	2 0819	2 1159	2 1011	2 0169	1 0702	1 0270	1 0854	1 0091
	2 0508	2 1451	2 1307	2 0619	2 0224	1 3353	1 0584	1 1438
	1 0993	1 1771	2 2034	1 1265	1 0723	2 0311	2 1802	2 1651
25	1 0456	1 0338	2 0560	2 0190	1 2028	2 1191	1 0195	1 1186
	2 0872	1 0013	2 0094	1 1448	2 0221	1 0436	1 0272	1 1935
	1 0290	2 0204	1 1218	2 0515	1 0693	2 1802	1 0223	2 0323
	1 0018	2 0368	1 0444	2 0107	2 1330	1 0271	2 0304	2 1589
	2 0944	2 0363	1 0869	2 0554	2 1213	1 0583	1 1224	1 0728
30	2 0029	2 0007	1 1513	1 0433	1 0929	2 0419	1 1935	1 1021
	2 1170	2 0339	1 0880	1 0614	2 2048	2 1237	2 0400	1 1178
	1 0187	1 0117	1 1225	1 0497	2 1129	2 0104	1 0664	2 0872
	2 0156	2 0316	1 1845	2 0267	2 0428	2 0220	1 0983	1 0808
	2 1439	2 1011	1 0132	1 0859	2 1145	1 0468	2 0136	1 1799
35	1 0577	1 1478	2 1799	2 1974	1 0085	2 0384	1 0460	2 0250
	2 0692	1 1000	2 2028	1 1718	2 0936	2 0976	1 0594	1 1449
	2 0356	2 1882	1 0398	1 0046	1 0030	1 1046	1 1587	1 0412
	1 0888	1 0174	1 0885	1 1704	2 0215	1 0875	2 1194	1 0698
	2 0417	1 0410	1 0477	2 0921	1 0744	1 0080	1 1163	2 0603
40	2 1576	2 0077	2 1586	2 0277	2 0816	1 0071	1 0947	1 0059
	2 0245	2 0112	1 1794	1 1018	1 0279	1 0948	1 1527	1 1672
	2 0723	2 0288	1 0080	1 0002	1 0107	2 2536	1 0794	1 1420
	1 0799	1 0038	1 0605	1 1328	2 1795	2 0612	1 0265	2 0005
	1 1840	1 0542	2 1556	1 0376	2 0219	2 2391	2 0824	2 1467
45	2 0701	1 1413	2 1653	2 0947	1 1255	1 0066	1 0237	2 0667
	1 0509	2 0874	2 0346	1 0338	2 0381	2 1776	1 0163	1 0088
	2 0923	2 0796	1 1036	1 2834	1 1840	1 0603	1 2428	1 0719
	2 1679	1 0381	2 0392	2 0900	1 0219	2 0359	2 0050	1 1154
	1 0981	1 0142	1 0593	1 0088	1 0209	2 0225	2 0976	2 0840
50	2 0046	2 0039	2 0791	2 0878	2 0214	1 1611	1 0955	2 0654

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

* For warning as to use see paragraph 2.10 of Introduction.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 8*

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0707	1 0703	1 0488	1 0750	1 0158	2 0517	1 0993	2 1200
	2 0690	1 0343	1 0227	1 0242	1 0719	1 1082	2 1190	2 1466
	1 0358	1 0014	1 0335	1 0442	1 0819	1 0054	1 1030	1 0219
	1 0452	2 0119	2 1530	2 1082	2 0069	1 0108	2 0419	1 0820
5	2 0715	1 0317	2 1524	1 0974	1 0422	2 1553	2 0114	2 1896
	2 1283	2 0562	1 0395	1 0673	1 1133	1 0833	2 0601	1 1177
	1 0771	1 0643	2 1730	2 0184	1 0832	1 0437	1 0608	2 1106
	1 0890	1 0209	2 1212	1 1255	2 0547	2 0775	2 1030	1 0992
	1 1309	1 0990	2 1855	2 1063	2 0577	1 0586	1 2468	2 0368
10	1 0228	2 1299	1 1107	1 0095	1 0648	2 1039	1 0568	1 1170
	1 0175	1 0261	1 0849	2 0376	2 1413	1 0801	1 0568	1 0182
	1 0514	1 1053	1 0319	1 0914	2 2748	2 0175	2 0357	2 0259
	2 0576	1 0120	2 1252	1 0740	1 1162	2 0938	1 1751	1 0845
	2 0235	2 0212	2 0880	1 0575	1 0564	2 1231	1 1404	1 0624
15	1 1218	2 0267	1 1888	2 0917	1 1179	2 0298	2 0770	2 0623
	2 1649	1 0958	2 0723	2 2727	2 0723	2 1310	1 0968	2 1900
	2 1049	2 0219	1 0510	1 1476	1 0586	1 0344	1 2206	2 0083
	1 1556	2 0086	1 0214	1 2562	2 2001	2 0287	1 1650	2 0940
	1 0152	2 0371	2 0329	2 0020	1 0740	1 2349	1 0388	1 0341
20	2 1624	1 0575	2 0861	2 0762	1 0195	2 0064	1 0743	2 2290
	1 0243	2 0425	1 0249	1 1131	1 0008	1 0421	2 0584	2 1187
	1 0436	1 0180	1 1138	2 1507	1 0444	2 0102	1 0018	1 1288
	2 0401	2 0118	1 0365	2 0973	2 1115	1 1181	2 1116	1 0124
	1 1054	2 0708	1 0069	1 0425	2 0789	1 1477	1 0283	2 0513
25	1 1410	2 1255	2 0812	1 0127	2 0509	2 0198	1 0557	2 0467
	2 0149	1 0910	2 1090	2 1436	2 0970	2 0612	2 0157	1 1547
	2 1887	1 1049	2 2118	2 0651	2 0990	1 1863	2 2183	2 1891
	1 0185	2 0580	2 1917	1 0464	1 0377	2 1887	2 1311	1 0788
	1 1062	1 0722	2 0064	1 0383	2 0470	1 0213	1 0739	2 1985
30	1 1495	2 0056	1 0361	2 0074	1 0527	1 0179	1 0435	2 1227
	2 0720	1 0252	2 0597	1 1189	2 2297	2 1111	1 1043	1 0967
	1 1002	1 1739	1 0984	2 0108	2 0403	1 1185	1 0083	2 0152
	1 0830	2 1189	2 1669	2 0478	1 0949	2 1155	1 0395	2 1094
	1 0626	1 1662	2 2370	1 0467	2 0980	2 0364	2 1478	2 0293
35	2 0402	1 0695	2 0810	2 0773	2 1276	2 1084	1 0296	2 0474
	1 0465	2 0516	1 0561	1 0277	1 0115	1 0574	1 1331	1 2028
	1 0226	2 0735	1 0562	1 0070	2 0783	1 0237	2 0731	2 0182
	1 0366	2 0208	2 0530	2 0725	2 0559	2 2644	2 1255	1 0094
	2 0347	1 0780	2 0164	2 0600	2 0542	1 0839	1 2357	2 2878
40	2 0911	2 1385	1 1117	1 0634	2 0240	1 0146	2 0531	2 0945
	1 1369	2 0726	1 1442	2 0170	2 0649	2 0589	2 0643	2 1617
	2 0697	2 0444	2 0434	2 1420	1 0944	2 0649	2 1211	1 1077
	1 0382	1 0177	2 0178	1 1336	2 0623	2 1041	1 0767	2 0536
	2 0096	1 1571	2 0714	1 1847	1 0982	1 2349	1 1406	2 0378
45	2 1650	2 0325	1 1699	2 1138	2 0687	1 1616	2 0656	1 1050
	1 0390	2 1723	2 0429	1 0492	2 0627	2 0409	1 1654	1 0588
	2 0679	1 1358	1 0280	2 0286	2 0199	1 0013	2 0720	2 0005
	2 0657	2 0621	2 0136	1 2697	1 1138	1 0795	1 0085	2 0219
	2 0647	2 0102	2 0216	2 0951	1 1662	1 0693	1 0768	1 0402
50	2 1403	1 1683	1 1967	1 0824	2 0361	2 1417	1 0545	2 1809

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

* For warning as to use see paragraph 2.10 of Introduction.

TABLES OF RANDOM NORMAL DEVIATES

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 9

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 2130	1 0308	2 0293	1 1088	1 0244	2 0643	2 0354	2 1354
	2 0359	2 1113	1 0289	1 1055	2 1038	1 0358	1 0562	2 0068
	2 0326	1 0667	1 1430	2 1069	1 1215	1 1866	1 1282	1 0996
	2 0157	1 2687	2 0069	1 1270	2 0002	1 1705	1 0255	1 0388
5	2 0315	1 0648	2 0506	1 0432	2 0300	2 0583	1 0344	2 0593
	1 1031	2 0211	2 1764	1 0197	2 0411	1 0563	2 0005	1 1577
	1 1739	1 1198	2 0603	2 0161	1 0056	1 0078	1 1006	1 0385
	2 0841	1 0060	1 0944	1 0241	1 0716	1 1062	1 1001	1 0308
	2 1678	1 1304	2 0661	1 0505	2 0384	2 0612	1 2270	2 1346
10	1 1266	2 1223	2 0533	1 0838	2 0337	1 2620	2 1587	1 0488
	2 0165	2 2073	2 0042	2 0719	1 0355	2 1030	2 0595	2 2142
	1 0449	1 1735	1 2442	1 0111	1 0288	1 0119	1 0356	2 0418
	1 1402	1 1875	1 0231	2 0419	1 0514	2 0999	1 0918	2 1999
	1 0231	2 0428	2 0558	1 0763	1 1004	2 1592	2 1061	2 1006
15	2 1815	1 0985	2 1020	2 0290	1 1171	2 0810	2 0317	1 0221
	2 1414	1 1567	1 0534	1 1171	2 0826	1 0071	2 0564	1 0543
	1 0973	2 1376	1 1407	2 2122	2 0344	2 1903	1 1470	2 1279
	1 0286	2 0383	1 0346	2 0154	1 0590	1 0844	1 0964	2 1270
	1 1309	1 0077	1 0374	2 1919	1 1355	2 0203	2 1540	2 0969
20	1 1080	2 0834	2 0663	1 0432	1 0022	2 0806	2 0680	2 0636
	1 0160	1 1060	1 0723	2 1188	2 0329	2 0797	1 0281	1 0264
	1 1169	2 0978	2 0869	2 0113	1 0348	2 0037	2 0543	1 0776
	2 1757	1 0188	1 1745	2 0530	1 0221	1 0894	2 1025	1 0781
	1 2409	1 0462	1 0293	2 0928	1 0820	1 1171	1 1239	2 0183
25	1 0843	1 2929	2 1664	2 0364	1 0318	1 0368	1 0645	2 0573
	1 0464	2 1276	2 1341	2 0160	1 0366	1 1068	1 2342	2 0749
	1 0360	1 1606	2 0499	1 0497	1 0795	1 0623	2 1000	1 0087
	2 0680	1 0997	1 1945	2 0139	2 2149	1 0637	2 2489	1 0656
	1 0124	2 0612	2 0748	1 0491	2 0621	1 0497	2 0242	2 0506
30	1 0450	1 0692	1 0750	1 0806	1 1155	2 2030	1 1251	2 0688
	1 1363	1 0230	2 2183	2 0033	2 0653	2 0456	2 0130	2 0817
	2 0543	2 2273	2 0207	1 0474	2 0707	2 1190	1 1237	2 1101
	2 0306	2 0341	2 0646	2 3432	2 0389	2 1196	2 1222	1 0501
	2 0525	1 0292	1 1058	1 0195	2 0657	2 0961	1 1301	1 0894
35	2 0752	2 0011	1 0415	1 1014	1 0822	1 0557	1 1606	1 0053
	2 0987	2 0064	2 1186	2 0280	1 1766	1 1127	2 1192	1 1096
	1 1025	1 0525	2 0187	2 0738	1 0378	2 0133	1 1341	2 1323
	2 0137	2 1818	2 2597	1 0502	2 0202	2 1800	2 1119	1 1849
	1 0954	2 0237	1 1447	2 0943	1 0489	1 0603	2 1847	1 0066
40	1 0281	2 0782	2 1983	2 0654	2 0083	2 0895	2 1671	1 1167
	1 1060	2 1104	1 1152	1 0393	1 0587	2 1102	1 2034	1 1478
	2 0980	2 1069	1 1386	2 0449	1 0308	1 0217	1 0383	1 1208
	1 0986	2 1490	2 0739	1 0345	2 0493	1 0103	2 1169	1 0952
	1 0996	1 0907	1 0973	2 1651	2 0766	1 0728	2 0339	2 1041
45	2 1691	2 0162	1 2040	1 1120	1 0037	1 0918	1 0858	2 0297
	1 0710	1 0072	2 0643	1 0435	1 0510	1 1052	1 0117	2 1206
	1 0005	1 0329	1 0326	2 0523	1 1299	2 0445	1 0256	1 1227
	2 1530	1 0114	2 1928	1 0618	2 1616	2 2113	2 1651	1 0267
	1 0166	1 0583	2 2948	1 0508	2 1134	2 0345	1 0733	2 0402
50	2 0417	2 0985	2 0545	1 0174	2 1023	2 0714	2 1464	1 1390

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 10

row no.	col. no. 1	2	3	4	5	6	7	8
	2 1232	1 1135	1 1687	2 0752	2 1770	2 0724	2 0094	1 1465
	1 1542	1 0807	1 2334	2 0944	1 0970	2 1224	1 0332	2 0091
	1 0520	1 1108	2 0281	1 0460	1 0176	2 0120	1 0772	2 2319
	2 2562	2 1278	2 0962	2 0659	1 1462	1 0019	1 0892	2 0346
5	1 0186	1 1439	1 0470	1 0051	2 0252	2 0331	2 1238	2 0895
	2 1036	1 0250	1 0650	2 0543	1 0848	2 1147	2 0965	2 0915
	1 0093	1 0309	2 0649	2 0496	1 0162	1 0564	2 0165	2 0813
	2 0081	2 0445	2 1455	1 0882	1 0364	1 0608	2 0314	1 0253
	1 0791	2 0655	1 1018	1 1267	2 0136	2 0924	1 0078	2 0575
10	2 0016	2 0613	2 0634	2 0189	1 0150	1 1101	2 0567	2 1165
	1 0617	2 1691	1 1279	2 1551	2 1653	2 0019	2 0256	1 0313
	1 0025	1 1018	1 0275	1 0431	1 0072	1 0253	1 0891	2 1240
	2 1446	1 0008	2 1377	2 0171	2 0677	2 0873	2 0603	1 0568
	1 0630	1 0949	2 0449	2 0091	2 0095	1 2652	1 0082	1 0650
15	2 1122	2 0433	2 0308	2 2238	1 0018	1 0898	2 0716	2 1363
	1 1139	2 0626	2 1025	2 1252	1 0049	2 0879	1 0518	1 1032
	2 1556	1 1611	2 0129	2 2312	1 0148	2 0941	2 3121	2 0309
	2 0401	2 0208	1 0400	2 0371	1 0639	2 0420	2 0341	1 0866
	2 1410	2 0551	2 0091	1 0273	2 1034	2 1112	1 0076	1 0273
20	1 0356	2 0506	2 1013	1 0670	1 0479	1 0856	1 0895	1 0566
	1 0606	1 0728	2 1068	1 0685	2 0592	2 0686	2 0071	1 0455
	2 0086	1 2248	2 3353	2 0275	1 1394	1 1480	1 0085	1 2737
	2 0026	1 1148	1 1551	1 1084	1 0549	2 1572	1 0612	2 0756
	1 0511	1 0458	1 1829	1 2178	1 0639	1 0800	2 0623	2 0360
25	1 0230	2 0464	2 0575	2 0296	1 1377	2 0198	1 1076	2 3719
	2 0448	2 1392	1 1460	2 1025	1 0991	2 1455	2 1902	2 1619
	1 0129	1 1271	1 1305	2 1266	1 0094	1 0879	1 0251	2 0005
	1 0631	1 0176	2 0054	2 0596	1 0791	1 0617	1 1337	2 0526
	1 0736	1 0725	2 0050	1 0038	1 1398	2 0490	1 0248	1 1779
30	1 0914	2 0336	1 0244	2 0392	2 0501	2 0550	2 0430	1 2197
	2 0779	2 0822	1 0815	2 1051	1 1057	1 1525	1 1395	2 1834
	2 0167	1 0799	2 1620	1 1496	1 0476	1 0124	1 0033	1 1082
	2 0526	2 0462	1 0502	2 1608	1 0290	2 0473	2 1416	1 1295
	1 0640	2 1080	1 1372	1 0810	2 1023	1 1161	1 1732	2 1187
35	2 0483	1 0346	2 0150	1 2052	1 0613	2 0191	1 0885	1 0423
	2 1204	1 0250	2 2489	1 0474	2 0498	1 0622	1 0534	1 0007
	1 0487	1 1597	2 1400	1 0875	2 1316	1 1613	2 0537	1 1096
	2 1725	2 1869	1 0220	1 0174	1 0726	1 1054	2 0625	2 1565
	2 1165	2 1410	2 1661	2 1914	1 1568	2 0666	1 0236	1 1520
40	2 1553	2 0020	1 0184	2 0835	1 1042	1 1394	1 0905	1 0237
	1 0973	2 1417	2 0288	2 1963	1 0187	2 1517	1 2020	2 0138
	2 1509	1 0243	2 1798	1 2687	1 0695	2 0137	1 0268	2 2024
	1 1382	1 0175	2 1030	1 0472	1 0346	2 0813	2 0013	1 0104
	2 0240	1 1307	1 0491	2 1016	1 0607	1 1623	1 1574	1 0204
45	2 0153	1 1957	2 0311	1 0702	1 0535	1 0547	2 0347	1 1997
	2 0181	1 1986	1 1040	2 0020	2 0732	2 0883	1 0814	2 0100
	1 0831	2 1262	2 0632	1 0053	1 1677	2 0412	1 0162	2 1341
	1 0452	1 0460	2 0588	1 0812	2 2155	1 0494	2 0571	2 0484
	2 0667	1 0216	2 1354	1 1710	2 1497	1 0345	1 1430	2 0508
50	2 0880	1 0846	1 2748	2 1930	2 1472	1 1755	1 0050	1 2007

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 11

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0242	2 0454	2 0569	1 0103	2 0402	2 0833	1 1182	2 0649
	1 0910	1 0506	2 0019	2 0955	1 0423	1 0483	2 1536	1 0567
	1 1281	1 0254	1 0962	1 0555	1 0421	1 0767	2 0460	1 0591
	1 0219	1 0646	2 1930	2 0664	2 0085	2 0521	1 1851	2 0928
5	1 1165	2 0042	1 1559	2 0166	1 0830	1 1092	1 2362	2 0019
	2 1972	2 1595	1 1900	2 0433	2 0858	1 0181	2 0737	2 1774
	1 0652	2 1453	2 0139	2 2878	1 0291	1 0460	1 1091	2 0951
	2 0099	1 1059	1 1708	2 0701	1 1169	2 0314	2 1151	1 0060
	1 1649	1 0695	2 0724	1 0294	1 1073	1 0482	2 0161	2 0197
10	1 0142	1 0350	1 0125	2 0264	2 0404	1 0079	1 0344	1 1029
	2 0532	2 0840	2 1328	1 0503	2 0014	2 0228	2 0349	2 1205
	1 0060	1 1813	2 0162	2 1351	1 1981	2 0911	1 1544	1 0113
	2 0066	1 0961	2 3540	1 1992	2 0753	1 2181	1 0935	1 0390
	2 0103	2 0261	2 0773	1 1352	1 1600	1 1611	2 0516	2 0177
15	2 0366	2 1368	2 0029	1 0753	2 0221	2 0235	1 0034	1 1164
	1 0067	1 0428	2 0428	2 0144	1 1034	2 0115	2 1301	2 1257
	1 0287	2 1193	2 0184	1 0992	2 0005	2 0803	2 1536	2 0144
	1 1997	2 0344	2 0771	1 0567	2 0946	2 1950	2 0660	1 1192
	2 0528	1 0643	1 0440	2 0165	1 0098	2 0570	2 0344	2 2197
20	1 0674	1 0932	1 0089	1 0534	1 0363	1 0497	2 1201	2 1125
	1 0325	2 0340	1 0882	2 0619	2 0227	1 0511	2 0625	1 0070
	2 0284	2 0635	1 1133	1 0284	2 0073	2 1300	1 1063	2 0464
	2 0212	1 2054	2 0532	2 1261	2 0276	2 0289	2 0562	1 0448
	2 0123	1 0382	1 1897	1 0717	2 1609	2 0549	1 1474	1 1203
25	2 0346	1 0192	1 0797	1 2326	1 2346	2 0215	2 0783	1 0066
	2 0471	1 0030	2 1509	2 1088	1 0199	2 0328	1 1612	1 1082
	1 1423	2 0263	1 0319	2 0157	2 0406	1 0209	1 0708	1 0317
	2 1160	1 0496	1 1095	1 1374	2 0118	1 0141	2 0904	1 0990
	1 0725	2 0339	1 0838	2 0581	2 0407	1 0080	1 0268	1 0792
30	2 2590	1 0188	1 0283	2 0930	1 0044	2 1231	2 0297	1 1178
	1 0260	1 0936	1 0301	1 0383	2 0119	1 1017	2 0009	2 0465
	1 0599	1 0079	1 2478	1 2052	1 0201	2 0706	1 0994	1 0506
	2 1140	1 0333	1 1158	2 1143	2 0241	1 1713	2 0693	2 1174
	1 0124	2 0210	1 1152	1 0255	2 2062	1 0908	1 0156	1 0196
35	1 0828	1 0088	2 0338	2 0693	1 0926	2 1711	1 0906	1 0594
	1 0096	1 2583	2 0437	1 0299	2 0299	2 0805	2 0861	2 1771
	2 0409	2 1899	1 0335	1 0114	2 2032	2 1146	1 0099	1 1377
	2 1507	1 0117	1 0147	1 1562	2 1560	1 1311	1 1089	1 0807
	1 0909	1 0656	2 1341	2 0548	2 0388	1 0881	2 0143	1 0821
40	1 0298	1 0324	2 0086	2 0555	1 0333	2 1502	1 1863	2 1188
	2 0522	2 0868	1 0452	1 0515	2 0176	2 0180	1 2315	2 1264
	2 0619	1 1086	1 0440	1 2194	1 0738	2 0750	1 1334	1 1297
	1 1876	2 0797	1 0321	2 1983	2 0975	1 1809	2 1360	2 0738
	2 1111	1 0608	1 1568	1 1621	1 1662	2 0873	1 1066	1 0636
45	2 0473	1 1437	1 0146	1 0247	2 1671	1 1235	2 0449	1 0249
	2 0868	1 1439	2 1218	2 0524	2 0301	2 2097	1 0064	2 0825
	1 0265	1 0934	2 0858	1 0124	2 1384	2 0528	2 1951	1 0689
	2 0822	2 1577	2 1210	1 1432	2 0733	2 0750	1 0095	1 1127
	1 1856	1 1278	1 1986	1 0159	1 0052	2 0587	1 1468	1 0707
50	2 1354	2 0796	1 0256	2 2848	1 1063	1 0211	2 1022	2 0249

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 12

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0087	2 0300	1 1007	2 0789	2 0722	2 0683	2 2628	2 0389
	1 1299	1 0712	2 1165	2 0220	1 0567	2 0862	2 0028	2 0935
	2 0831	2 0730	1 0629	1 0255	1 1563	1 1391	2 0606	1 1013
	2 1521	2 1041	2 1386	2 1768	1 1031	2 1726	1 0586	2 0013
5	1 1027	1 0269	1 0149	1 0338	1 1588	2 0001	2 1116	2 0021
	1 2395	2 0388	1 0134	2 0650	2 0164	2 1298	2 0650	2 0862
	1 0352	1 2323	1 1446	1 0029	2 1110	2 2181	2 1491	2 0240
	1 0609	2 1291	2 0458	2 0453	2 1916	1 0842	1 0268	2 0125
	2 1562	2 0022	2 0397	1 1162	1 0037	2 0669	2 0271	1 0466
10	1 0385	1 1352	2 0329	1 1194	2 1483	2 1069	1 0761	1 1689
	2 1426	2 1222	1 0810	2 0119	1 1161	2 0887	2 0802	2 0055
	1 1116	2 0032	2 0820	2 0902	2 1308	1 0007	2 0534	1 0402
	2 2101	1 0770	1 0382	1 0632	2 0229	2 0975	2 0744	1 0756
	2 0415	2 1303	1 0349	1 0192	2 1849	2 0646	2 0398	1 1283
15	2 1937	1 0256	2 1121	1 0221	1 0234	1 1079	1 1311	1 1279
	2 0230	1 0749	1 0901	2 0583	2 0847	2 1130	2 2644	2 0674
	1 1535	1 0633	1 0837	2 0864	2 0755	2 0409	1 0450	2 1790
	2 0513	2 0311	2 0794	1 0720	2 0246	2 0993	2 0669	1 1413
	1 1265	1 1104	2 0895	1 0071	1 0159	2 0628	1 0439	1 0119
20	2 1783	2 0214	1 0050	1 0193	2 0818	1 0420	2 0496	1 0180
	1 1336	2 0057	2 3540	1 0524	1 0772	2 0741	2 0544	1 0286
	2 0686	2 1260	2 0478	2 0367	1 0570	2 0831	2 0663	1 0548
	2 0517	1 0309	2 0489	1 0216	1 1688	2 0342	2 1951	1 0137
	1 0190	1 1437	2 1193	2 0445	2 0891	1 0069	1 0118	2 1105
25	2 0162	2 0328	1 0510	1 1037	2 1006	1 0192	1 0012	1 0564
	1 0576	1 0964	1 0986	1 0633	2 0106	2 0034	2 1531	1 0025
	1 0009	1 0238	2 0341	2 0786	1 0440	1 0597	2 0939	2 0606
	2 0435	2 2014	2 0356	1 2280	2 0098	1 1110	2 0777	1 0610
	1 1315	2 0940	1 0654	1 0474	1 0418	1 0949	1 1227	1 0134
30	1 0802	1 0046	1 0705	2 1873	2 0024	2 0438	1 0873	2 0280
	2 0425	1 1017	1 0639	2 1111	1 0087	1 0403	2 0479	2 0256
	2 0408	1 0729	2 1375	1 0553	2 0919	1 1597	1 0328	2 0667
	2 0275	1 0220	1 0806	1 0286	2 0749	2 0879	2 1436	1 0262
	1 0289	1 1308	1 0510	1 1197	1 0394	2 1330	1 0432	1 0182
35	2 0299	2 0953	2 0572	2 0081	2 0091	1 0622	2 0480	2 0442
	2 1104	2 1731	1 1484	1 2028	2 1134	2 1426	1 1004	2 1239
	2 0193	2 1963	1 0759	1 0386	2 0597	2 0242	2 1377	1 1329
	1 0490	2 0467	1 0537	2 0343	1 0066	2 1782	1 0694	1 0412
	2 1465	1 0002	1 0019	2 1209	1 0086	2 0991	1 0045	2 1301
40	2 0399	1 0364	2 0439	2 1569	2 1146	2 1505	1 0600	2 1428
	1 0179	1 0626	1 0687	2 0336	1 1552	1 0084	2 0245	2 1042
	2 0647	2 0440	2 1596	2 0698	2 0359	2 0650	1 0529	1 1389
	1 1072	2 0188	1 0645	1 0066	1 0103	2 0404	2 0498	1 1068
	1 1026	1 0338	2 0267	1 1845	1 0394	2 1209	1 1163	2 0767
45	2 1924	1 0157	1 1058	1 0139	1 0665	2 0583	2 0576	1 1774
	1 0638	1 0771	1 0820	2 0627	2 0137	1 0522	2 0539	1 1068
	1 0032	2 0810	1 0748	2 1314	2 0279	2 0140	1 0558	1 0910
	2 0819	2 0316	2 1317	1 0619	1 0238	2 0404	2 0831	2 0062
	1 1487	1 0910	1 0774	1 0330	1 0545	1 0383	2 0250	2 1272
50	1 0472	1 1151	2 0396	1 0254	2 0439	2 0281	1 0634	2 0404

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE NO. 13

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0143	2 1965	1 1955	2 1737	2 0281	2 0674	2 0532	2 0214
	1 1400	1 0617	2 1036	2 0075	1 0533	1 1752	2 1530	2 0325
	2 0190	2 1481	1 0264	2 1322	2 0538	2 0008	1 0930	1 0748
	1 0517	2 1253	1 1262	1 0039	2 0788	1 1704	1 0721	2 0512
5	1 0170	2 0994	1 1146	2 1632	1 1133	1 0281	1 0622	1 0738
	1 0634	1 0619	2 0227	1 0260	2 0200	1 0011	2 0125	1 0780
	1 0506	2 2132	2 0435	1 0311	2 0220	1 0392	2 0171	1 1574
	2 0998	1 0511	1 0795	1 0221	2 1402	2 1207	1 1579	2 2077
	2 0080	1 0597	2 0842	2 1307	1 0964	2 0253	2 0752	1 1197
10	2 0200	2 0441	2 0942	1 0462	1 0998	1 0512	1 0073	1 0010
	1 0224	2 0435	2 1182	2 1167	2 0276	2 1223	1 0589	2 2697
	2 0648	2 0196	1 1587	1 1815	2 1828	1 1899	1 1134	1 1736
	2 1138	1 0247	2 0183	2 0026	1 0189	2 0492	1 0021	1 0266
	1 0468	2 0951	2 2669	2 1160	1 0056	2 0683	1 1231	1 2748
15	1 0483	1 1463	2 1177	2 1816	2 0486	1 0228	2 0711	1 0817
	2 0465	2 0165	2 0139	2 1997	2 1785	1 0942	2 0139	2 0427
	2 0961	1 0187	2 0879	2 0674	2 0972	1 0177	2 0097	2 0225
	2 0709	2 0554	2 0411	2 0039	1 1153	1 0846	2 1967	2 0342
	2 0661	2 0276	2 1067	1 0353	2 0421	1 1305	1 0249	2 1710
20	1 0466	1 0250	1 0898	2 1426	2 0427	2 0250	1 0526	1 1292
	1 0885	2 0089	1 1836	2 0410	1 1215	2 0376	1 0138	1 0504
	2 0008	2 0608	1 2530	2 1208	1 1649	2 1627	1 0060	2 1287
	2 0127	2 1222	1 0930	1 1387	1 0317	1 0592	2 1287	2 1052
	1 1294	1 0013	1 0096	1 1730	2 1114	2 0669	1 0148	2 0240
25	2 0046	2 1278	2 1244	2 0563	1 0812	2 0621	1 0145	2 0964
	1 1225	2 0033	2 0462	2 0449	2 0367	2 0422	2 0338	2 0774
	2 0193	1 0672	1 0913	1 0018	1 0400	2 1406	2 0863	2 1159
	1 0505	1 0198	1 0433	1 0790	1 0964	1 1562	2 0453	1 0467
	2 1643	2 0350	1 0770	2 1024	2 0236	1 0115	1 1099	2 0455
30	1 0369	1 1598	1 1224	2 1062	1 0745	2 0038	2 0470	1 0242
	2 1151	2 1745	2 0107	1 0485	2 0322	1 1704	2 1267	1 1130
	2 0493	2 1191	1 0265	1 0211	1 0317	1 0452	2 0719	1 0742
	1 1186	2 0357	1 0787	2 0812	1 0814	1 0834	2 1516	1 1283
	1 1117	2 1174	2 0663	2 1869	1 1142	1 1914	2 0112	2 0079
35	1 0082	2 1150	1 1647	2 1825	1 2304	2 1708	2 0620	1 1405
	2 0306	2 1402	1 0855	1 0666	2 1630	1 1134	1 0994	1 0139
	2 0616	1 0707	2 0410	1 0392	2 0610	2 0114	1 1376	2 0985
	1 1059	1 0973	1 0818	2 1024	2 0379	1 0491	2 0168	1 0594
	1 1133	1 0161	1 0713	2 0482	2 0124	2 0489	2 0247	2 0111
40	2 0805	2 0071	1 0701	2 0106	1 0880	1 0683	1 1330	2 0698
	1 0386	1 1036	2 0496	2 0769	2 0411	1 0336	1 0499	1 0860
	2 2387	2 1468	1 1703	1 1006	2 0168	1 1859	1 0383	1 0650
	2 0300	2 0411	1 0152	1 0108	2 2395	1 0546	1 0504	1 0540
	2 0775	1 1394	2 0046	2 1584	2 1168	1 1029	2 0745	2 1331
45	1 0098	2 1085	1 0553	2 0106	1 0027	1 0822	2 0125	2 0593
	2 0092	1 0297	2 0500	2 0177	1 1089	2 1333	1 0501	1 0091
	2 1650	2 0328	1 0267	2 0945	2 2277	1 0937	2 0088	1 2330
	2 0441	2 0505	2 0513	2 1371	2 0542	2 0249	1 0504	1 0742
	2 1104	1 0144	1 0604	1 0479	1 0659	1 0236	1 0378	2 0319
50	2 0487	2 0642	1 0144	2 1024	1 0879	1 0131	1 0606	1 0820

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 14

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0778	1 0405	2 1404	1 0336	2 1474	2 0159	1 1924	2 2058
	2 0467	1 1435	2 0016	1 0915	2 0352	2 0033	1 0507	1 0669
	2 0531	2 0409	2 0571	2 0191	2 0699	2 0847	1 1486	2 0732
	1 0955	1 2101	1 0224	2 0703	2 1196	2 0604	1 0370	2 0541
5	1 0140	1 1048	1 0297	1 0927	2 0694	1 1017	2 0096	2 0140
	2 0633	1 0806	2 0421	1 1379	2 0986	2 0284	1 0758	2 0013
	2 0465	2 0020	1 0615	2 0021	1 0731	1 1142	1 2794	1 0449
	2 0598	1 0305	1 0572	2 0215	1 0549	2 0557	1 1045	2 0986
	1 0667	2 2287	2 0982	2 0705	2 0530	1 0838	1 1856	1 0031
10	1 0095	1 0546	2 0615	2 0546	2 0151	1 0130	1 0880	1 1037
	1 1449	2 0099	2 0837	2 0274	1 0761	2 0301	2 0452	2 0978
	1 0217	1 0761	2 0476	2 1645	2 0646	2 1603	1 0093	1 0777
	2 0557	1 0420	1 0456	1 0927	2 0999	2 2501	2 2597	2 0302
	1 0033	1 0910	2 0329	2 0120	1 0430	1 0082	1 0096	2 0612
15	1 1314	1 0582	2 0782	2 0449	2 1219	2 1088	2 0161	2 0129
	1 0806	1 0220	2 0641	1 1182	1 0128	2 0170	2 2748	2 0063
	2 0334	1 1413	2 1592	1 0634	1 0572	1 1715	2 1121	1 1587
	2 1774	2 0429	2 2203	2 0111	2 0201	1 0384	2 0304	2 1816
	2 1967	2 0133	1 1909	1 0961	2 1134	2 1803	2 0710	2 0710
20	1 0603	2 2404	1 0495	2 0568	2 0084	2 1538	2 0101	2 0412
	1 1485	2 0055	1 0186	1 1236	2 0113	2 0255	1 0144	1 1425
	1 0282	1 0279	2 1302	1 0629	1 0481	2 0515	1 2834	2 0313
	2 0736	2 0132	1 1371	2 1492	2 0624	2 0820	2 1624	1 1428
	1 2111	1 1204	1 1115	1 0424	1 0118	1 1185	2 0174	2 0805
25	1 0441	1 0934	1 0270	1 0514	2 2556	1 0911	1 0384	2 0123
	2 1221	1 1631	1 0394	2 0265	2 0862	2 1030	2 0961	2 0411
	1 1813	2 1055	1 0154	2 1803	2 0191	2 0429	1 1757	2 0367
	1 0974	1 0184	2 0167	2 1389	1 0806	2 1123	1 0205	1 1121
	1 0519	1 0032	1 0751	2 0417	1 0309	2 0401	1 1070	2 0746
30	1 0383	1 0721	1 0549	1 0979	1 0167	2 0217	1 2248	2 0319
	1 1067	2 0789	1 0893	2 0162	1 0081	2 0659	2 1911	1 0688
	1 1385	2 1741	2 0163	1 0285	1 1233	1 2457	2 0210	1 0262
	2 0797	2 2139	1 0299	2 0008	1 0910	1 1485	1 1149	1 1069
	1 0874	1 2232	2 0853	2 1938	2 1490	2 0624	1 0476	2 0494
35	2 0536	1 1288	2 1059	2 0165	1 0530	1 0896	2 1012	1 0994
	2 1552	2 0843	2 0483	1 0511	2 1109	1 0677	2 0415	1 1389
	2 1426	2 2489	2 0010	1 0392	2 0116	1 1554	1 0017	2 2022
	1 0594	2 1174	2 0522	2 0169	1 0076	1 1449	2 0558	2 0050
	2 0966	2 0549	1 0507	2 1480	1 2512	1 0693	2 0058	2 1702
40	2 0399	2 0373	1 0632	2 0067	1 0235	2 1176	1 0087	1 0962
	2 0719	1 2378	1 1171	1 0636	1 1595	2 2495	1 2058	2 0973
	2 0968	1 1504	2 0775	2 0895	1 0438	2 0230	1 0062	1 0479
	2 1208	1 0255	2 0320	2 1073	2 0121	1 0863	2 2157	2 0178
	1 0125	2 0629	1 0145	2 0025	2 1807	2 0335	2 2834	2 0105
45	2 1014	2 1202	1 0445	2 0015	1 1115	2 0161	1 0114	2 1080
	1 1142	2 0369	2 1729	1 0148	1 2697	1 1231	1 0387	2 0824
	1 1298	1 2418	1 1152	1 0773	2 2287	2 0160	1 0634	2 0322
	2 0189	2 0027	2 0916	1 1173	2 0210	2 1663	2 1549	2 0123
	2 0902	1 0023	1 0600	2 0043	1 0437	1 1455	2 0869	2 0371
50	2 0714	2 1571	2 0129	1 0708	2 1645	1 0163	2 0691	1 1363

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

TABLES OF RANDOM NORMAL DEVIATES

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 15

row no.	col. no. 1	2	3	4	5	6	7	8
	1 0711	1 0364	1 1361	2 0217	1 0537	2 1101	1 2086	1 0001
	2 1705	2 0076	2 0427	1 0043	2 0190	2 0545	2 0530	1 1021
	2 0414	1 1093	2 0137	2 0158	1 0928	1 0197	2 0627	2 0526
	1 1258	2 0221	2 1449	2 0328	2 1221	1 0101	1 2612	1 1396
5	1 0742	2 1001	1 1070	1 2071	2 0295	1 0579	2 1257	2 1765
	1 0042	2 0138	1 0766	2 0115	1 0461	2 0048	2 0328	2 0923
	2 1561	1 0066	1 0707	1 0863	2 0383	2 0290	1 0348	2 1141
	1 0422	2 0591	1 1367	2 0586	2 0335	2 0203	2 1386	1 0570
	2 2001	1 1010	2 0013	2 1299	2 0214	2 0494	1 0902	2 1301
10	1 1851	2 0100	1 0334	1 0209	2 0771	1 0227	1 1232	2 0388
	2 0358	1 1937	2 0322	1 0160	1 1515	1 1217	2 1074	1 0972
	2 1626	2 0903	1 0193	1 0860	2 0301	1 1941	2 0103	2 1457
	2 0173	1 1595	2 0761	2 0443	1 2391	2 0554	1 1365	2 1600
	2 0027	1 0727	1 0154	1 1653	1 0204	1 0391	1 0465	2 0851
15	2 0957	1 1296	1 0044	2 0902	2 0618	1 0335	1 0031	2 0422
	1 0792	1 1740	2 0886	1 0494	2 0369	1 1270	2 0196	2 0071
	1 0365	1 0352	1 0805	2 0900	2 0361	1 0095	2 1228	2 0150
	1 1999	2 0223	1 0175	1 0294	2 1054	2 0174	1 0163	2 0033
	2 1086	2 0272	1 1783	1 0714	1 0758	1 1209	1 0214	1 0494
20	1 1031	2 1604	1 0986	1 0032	1 0088	1 1795	1 0297	2 0824
	2 0949	1 1649	1 0258	1 0025	2 0893	2 0454	2 0506	1 1790
	2 0062	2 0192	1 0518	2 1233	2 0625	2 1957	2 0561	1 2111
	2 0156	2 1255	2 1084	2 0602	1 0135	1 0166	2 0571	2 0580
	2 1282	1 1190	2 1174	1 0386	1 1016	2 1355	2 1856	1 1196
25	1 0256	2 0210	2 1007	2 1498	2 2095	2 1039	2 0378	1 1171
	1 1689	2 0293	1 0747	1 0213	1 0401	1 0156	2 0679	2 2127
	2 0917	1 0350	1 1059	1 0121	1 0090	1 1582	1 2400	1 0969
	2 0336	2 1581	2 1458	1 1036	1 1252	2 0821	1 0314	1 0062
	2 0033	2 1238	1 0590	2 0014	1 0566	2 1045	2 0406	1 0510
30	2 1584	1 0579	1 1523	1 1130	1 2040	1 1110	1 0100	2 0141
	2 0134	2 0667	1 1254	1 2127	1 1087	2 0505	2 1052	2 1186
	2 1460	2 2597	2 0336	1 0626	2 0684	2 0056	1 0088	2 0467
	2 2147	2 1451	1 0655	2 2257	1 0630	1 0053	1 1378	1 0860
	2 0230	2 1383	1 0265	2 0846	2 0178	2 2090	2 0547	1 0148
35	2 2214	1 0791	1 0285	2 0167	2 0143	2 1454	1 0659	2 0600
	1 1031	1 0427	2 0778	1 1351	2 0909	2 1354	2 0219	1 0631
	1 0099	2 0143	1 1109	2 0011	1 0402	2 1387	2 1251	2 0655
	1 0277	2 0052	2 0234	1 0106	1 1084	1 0231	2 0506	1 0514
	1 0353	2 0775	1 0709	1 1060	1 0616	2 1820	1 0151	1 0087
40	2 0127	2 1203	1 0131	1 0919	2 0555	2 2068	2 0822	2 0057
	2 0517	1 0093	2 0821	1 0025	1 0981	1 0553	1 0546	1 0554
	1 0482	1 0628	1 1181	2 1217	1 0624	1 2506	2 0908	2 0393
	1 0788	1 1699	2 0570	2 0360	2 1179	2 1183	2 0571	2 1055
	1 0312	1 0415	1 0466	2 0689	1 1927	2 0360	2 0546	1 0283
45	2 0634	2 0416	1 0125	1 0213	2 0871	2 0097	2 0528	1 1750
	2 0148	1 0292	2 0722	1 2530	2 0115	2 1301	1 0002	2 0763
	1 1072	1 1637	2 0371	2 0205	1 0458	2 1044	1 1197	2 0296
	1 0082	2 0509	1 2162	2 0638	1 0298	1 0942	1 1037	1 1154
	2 1492	2 0895	1 1227	2 0187	1 1784	1 0530	2 0584	1 0913
50	2 0067	1 0944	1 1336	2 0463	1 0403	2 0593	1 0561	1 0038

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE NO. 16

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0875	1 1719	1 2462	2 0299	1 1320	2 3432	2 0778	2 0253
	1 0114	1 0474	2 0828	1 0193	2 1734	1 0656	2 0370	1 0694
	2 0384	1 2200	2 0230	2 0704	1 0497	2 0002	2 0201	2 0278
	1 0166	0000	1 0592	1 1010	1 0023	2 0057	1 2290	1 0310
5	2 0902	1 1642	2 0156	2 0678	1 1664	1 1521	1 0974	2 0085
	1 0617	2 0041	1 0756	1 0975	2 0257	2 0439	2 0328	2 0964
	1 0222	2 0367	1 1164	1 0814	1 1953	2 0309	2 0197	1 0840
	1 1165	2 0974	1 0501	1 1269	2 0115	2 0728	2 1568	1 1232
	1 0942	2 0172	2 1562	2 0281	2 0898	1 1792	2 1363	1 0623
10	1 1872	2 0804	1 1601	1 0289	2 1461	2 0558	1 0595	2 0553
	1 0021	2 1390	1 0104	2 2245	2 0151	1 0944	1 0576	1 1042
	1 2391	1 0831	1 1301	1 0664	1 1184	1 0811	1 0233	1 2086
	2 1001	2 0954	2 0050	1 0888	2 1170	1 1252	1 1662	2 0647
	2 0266	1 0589	1 0818	2 0773	1 0426	2 0544	2 1863	2 1503
15	2 0975	1 0037	1 1090	1 0083	2 2342	1 0028	2 0716	2 1351
	1 1144	1 0750	2 1776	2 0816	2 0619	1 0938	1 0325	1 0568
	1 0484	2 0106	1 0236	1 1730	1 2028	1 1510	1 0284	1 0501
	2 0124	1 0890	1 0185	2 1606	1 0467	2 0586	2 0984	2 0170
	2 0804	2 1164	1 0439	2 0931	2 1587	1 0879	1 0014	2 0233
20	2 0669	2 0982	1 0217	2 1112	2 0619	1 1468	1 0988	1 0066
	2 0432	2 0527	2 1081	1 0659	1 0299	1 1138	2 0124	2 1460
	2 0368	2 0742	2 1719	2 0827	1 0410	1 1375	1 0599	1 1204
	1 1483	1 0163	2 1490	1 1243	1 0663	2 0264	2 0095	1 0114
	1 0603	1 0788	2 0149	1 1472	2 1594	1 1587	1 1569	1 0482
25	1 2506	1 0676	2 0428	2 1004	2 1616	1 0940	2 0253	2 0546
	1 1215	2 0465	2 1419	1 0429	2 1916	2 1283	2 1715	1 1249
	1 1251	1 2022	1 0438	2 0677	1 0260	1 0568	1 0537	2 1360
	1 0428	1 0081	2 0704	2 0010	2 1843	1 0132	2 0284	1 0031
	1 0180	1 0411	2 0404	2 0668	2 0770	2 0269	1 0793	2 0883
30	2 1161	1 0750	2 0445	1 1031	1 1336	1 1331	1 0614	2 1427
	2 0746	2 0055	1 0012	2 0007	2 0322	2 1273	2 0008	2 0638
	2 0504	2 1198	1 0333	1 0132	1 0273	1 0400	2 0096	1 1086
	2 1232	1 2301	2 0455	2 0092	1 2101	2 0899	2 0101	2 0570
	1 0323	2 0533	2 0768	2 0713	2 0174	2 0191	1 0775	1 0861
35	2 1014	1 0028	2 2620	1 2264	1 0383	1 0947	2 1305	2 0554
	2 0900	2 0209	2 0990	1 0032	2 0281	1 0293	1 0035	2 1255
	2 0014	1 0191	1 0081	1 0089	1 0433	1 2001	1 0862	1 1360
	2 1616	2 1295	1 0020	1 2569	1 0006	2 0217	1 0512	1 1549
	2 0893	2 0714	2 1400	2 0110	1 0059	2 0021	1 1081	2 0722
40	2 0909	2 0129	1 1256	1 0033	1 0207	1 0167	1 0982	2 2346
	1 0871	1 0743	1 0363	2 0165	1 0499	1 0755	2 1406	1 0324
	2 0519	2 0083	2 0095	2 0726	2 2323	1 0886	2 0277	2 0167
	2 0877	1 0820	2 0230	2 0402	2 1134	2 1639	1 0218	2 0408
	1 0178	2 1625	1 0543	2 0202	2 2357	2 1818	1 0846	1 0348
45	2 0291	2 1849	2 0798	2 0444	2 0121	2 0123	1 1427	2 1248
	1 0263	1 0666	1 0206	2 1953	1 0139	1 0287	2 0199	2 0206
	2 1049	1 1281	1 0925	1 2697	1 0588	2 0528	2 0432	1 0429
	1 1101	1 0129	2 0476	2 0073	2 2468	2 0453	2 1865	2 1295
	1 0502	2 0776	1 2644	2 1122	1 0085	1 1034	2 0509	2 1350
50	2 0500	1 0705	2 0908	2 0865	1 0189	1 0575	2 0814	1 1607

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 17*

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0255	2 0410	1 0185	2 0928	2 1653	2 1702	1 1158	1 0753
	1 2948	1 2576	2 1123	2 0079	2 1128	1 1113	1 1717	1 0139
	1 0294	2 0359	1 0146	2 2014	2 1270	2 1460	1 0922	1 0191
	2 1393	2 0537	2 0326	2 0879	2 0568	1 0510	1 0431	2 0147
5	2 0177	2 0953	2 0443	2 1553	2 0614	2 0478	1 0006	1 0654
	2 0163	1 0406	1 1027	1 0875	1 0522	2 0701	2 1896	1 0194
	2 1125	2 0438	1 1188	1 0237	1 0501	1 0887	2 0299	1 0649
	2 0438	2 0851	2 1498	2 0261	2 1330	1 1003	2 0147	2 0639
	1 0776	2 1116	2 0040	2 0315	2 0390	2 0727	2 1566	1 2273
10	1 1066	2 0478	2 0516	1 0551	1 0134	2 0553	2 1768	1 2414
	1 1052	1 0265	2 1488	1 0336	1 1171	2 0339	2 1583	1 0818
	2 1748	2 0652	2 0262	2 1740	2 0392	1 0754	1 0362	2 0079
	2 0528	2 0116	2 1440	2 0161	2 0683	2 0404	2 0468	2 0202
	1 1317	1 0314	1 0781	2 1817	2 1483	2 0998	1 0412	2 1280
15	2 0268	1 1266	2 0659	2 0410	2 1397	1 0528	1 1432	1 0659
	2 0819	1 0823	2 2028	1 0261	2 0424	2 0890	1 0329	2 0915
	1 0020	1 0272	2 1766	2 0768	1 0453	2 0333	1 0901	1 0995
	2 2206	2 2178	2 0900	1 0442	2 1055	2 0395	2 1933	2 1972
	2 0415	2 0054	2 0603	2 0775	2 1074	2 0312	2 0694	1 1466
20	2 0474	1 0088	2 0186	2 0542	2 1478	1 1563	2 0522	2 1471
	1 1596	1 1114	2 0059	1 0759	2 1387	1 0020	1 0089	2 0388
	1 2543	1 0395	1 0341	2 0114	1 0979	1 1198	1 1223	2 1089
	1 2301	1 2848	2 0346	1 0212	1 0468	1 1741	2 2111	1 1176
	2 1605	2 0051	1 2147	1 0286	1 0294	2 1430	1 0613	1 1752
25	2 1185	2 0313	1 0183	1 0506	1 0410	1 2028	1 1244	1 1322
	2 0773	1 2075	2 0669	1 0796	1 1302	2 0181	2 1871	1 0680
	2 0726	2 1180	1 1011	1 0399	2 1171	2 0557	1 2086	2 0329
	1 0294	1 1771	2 0840	2 0678	2 1693	1 0427	1 1187	1 1050
	1 1233	2 1104	1 0952	1 1487	1 0225	1 0519	2 0217	2 0992
30	2 1075	2 0469	1 1579	2 0440	2 3062	1 1202	2 0322	2 0765
	2 0745	1 0495	2 0664	2 0443	2 0217	2 1702	2 0424	1 0335
	2 0807	2 0914	2 2495	1 0784	2 1134	2 0692	1 0351	1 0043
	2 0718	1 0505	1 0774	1 1106	2 1771	1 1215	1 2064	1 0327
	2 1171	2 1943	2 0429	2 0053	1 0998	1 1361	2 0947	1 0332
35	2 0718	1 0172	2 0199	2 0191	2 0841	1 0349	2 1148	1 0947
	2 0071	1 0194	2 0070	1 0700	2 1216	2 1004	1 1140	2 1003
	1 0051	2 0254	1 1360	2 0345	2 0124	1 1422	1 0650	1 1847
	2 0632	1 0150	1 1349	1 1006	1 0868	2 0940	2 0827	1 0866
	1 1859	1 0186	1 0161	1 1317	1 0415	1 1143	1 0346	2 0894
40	2 0908	2 0432	1 0362	2 1146	2 0469	2 0582	2 0399	2 0317
	2 0269	2 0093	1 1086	2 1499	1 0336	1 2183	2 0426	1 0470
	1 0639	2 1970	2 0821	1 0222	2 1462	1 1073	1 0261	2 1494
	2 0816	2 0778	1 1093	2 0093	1 0249	1 0601	1 2462	1 0289
	1 1402	1 0280	1 0280	2 1004	2 0319	1 1428	2 2052	2 0226
45	1 0184	1 1036	1 0189	2 0789	1 0257	1 1466	2 0702	2 0325
	1 0070	1 1105	1 0245	2 0259	1 1562	1 1574	2 0107	1 0939
	2 0432	1 2378	2 2095	2 0872	1 1836	2 0906	1 0466	2 0803
	1 1083	1 0758	2 0197	1 0586	1 0296	1 1423	1 0840	2 0595
	2 0898	2 0158	1 1610	1 1524	2 0602	1 2727	1 0587	2 1297
50	1 1077	2 0902	1 0883	2 1040	2 0365	2 0187	2 0137	2 0165

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

* For warning as to use see paragraph 2.10 of Introduction.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 18

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0503	1 0479	2 0245	2 0247	1 2127	2 1030	2 0537	2 0929
	2 0152	1 0511	1 1582	1 1245	2 0553	1 0161	2 0833	2 2848
	1 2315	2 0871	1 0622	2 0154	2 1189	1 1532	2 1828	2 0376
	1 0831	2 3291	1 1793	2 1905	2 0080	1 0211	2 0798	2 0747
5	2 1470	1 1136	1 0966	2 0373	1 0118	2 1920	1 1865	1 1687
	2 1413	1 0167	2 0673	2 0116	1 0747	2 0586	1 0565	1 0751
	2 0518	1 0920	2 1532	2 0301	2 0210	2 1847	2 1483	2 0481
	1 0289	2 0058	1 0673	2 0686	1 0168	2 0278	2 0219	2 0992
	1 0979	2 1562	1 1061	1 0385	1 0886	2 0460	2 0777	2 0210
10	2 2678	1 0539	2 0890	2 0010	1 0278	2 0485	1 0019	2 1793
	2 1069	1 1261	1 1518	1 0400	1 0827	1 0475	1 1320	1 0254
	1 1584	2 0785	1 0078	2 0296	1 0635	1 1542	1 0207	1 0102
	1 0496	1 0350	2 0209	1 0483	2 1416	1 0545	2 1231	2 0935
	1 0167	2 1128	1 1009	2 0367	2 0612	1 0645	2 0345	1 0041
15	2 0322	1 0319	1 0415	2 0880	1 0520	2 0886	1 0831	1 0172
	2 1516	1 0413	2 0295	1 0023	2 0856	1 0970	1 0510	1 0475
	2 1737	1 1840	1 0171	1 1557	1 0490	2 0073	2 0657	2 1285
	2 0287	2 1974	2 0358	1 0432	2 0432	1 0323	2 0689	2 1132
	1 0660	1 1460	1 0174	2 1296	1 1066	1 0185	1 0753	1 0027
20	2 0387	2 1444	2 0727	1 0942	2 0714	1 1495	2 0129	2 0592
	2 1011	1 0630	1 0278	2 0433	1 1543	2 1211	1 0210	1 1276
	2 0469	1 1136	1 0382	1 0321	1 0327	1 1112	2 1536	2 1725
	2 0285	1 0583	1 0968	2 0042	1 1776	1 0854	2 0275	2 0091
	2 0768	2 0071	2 0231	2 0085	1 0667	2 0694	1 0828	2 0880
25	2 0739	2 1225	1 1255	1 0313	2 0990	1 1674	1 0055	1 1174
	1 0342	2 0007	1 1386	2 2257	1 0630	1 0777	1 0113	2 0065
	2 0248	2 0492	1 2495	2 0647	1 1780	1 0931	1 0613	2 1407
	2 0601	2 1231	1 0271	1 1768	2 0379	2 0123	2 0809	2 0783
	1 0408	1 0468	1 1261	1 0137	2 2149	1 0187	2 0837	2 1219
30	1 0198	2 1767	2 1752	1 0668	2 1389	2 0129	2 1890	2 0746
	2 2301	1 0994	2 2084	1 0293	2 1903	2 0381	1 0629	1 0085
	2 0224	1 1589	2 0026	1 0573	1 0494	2 0736	1 0299	2 1455
	1 0815	1 0860	1 2669	1 1141	2 0956	2 0701	1 0302	2 0614
	2 0280	1 0023	2 0633	2 1845	1 0613	2 1021	1 1284	2 1373
35	2 0860	2 0062	1 0247	2 0647	1 1770	2 1110	1 0735	1 0767
	1 0027	2 0218	2 0552	1 0061	1 0837	2 0389	2 0186	2 0811
	2 0440	1 0119	2 1065	2 1193	1 0044	2 0690	2 0261	1 0747
	1 0346	2 1537	2 0383	2 1103	2 0682	1 0972	2 0837	1 1091
	2 0428	2 1655	2 0092	1 0531	1 0167	2 1102	2 0475	1 0800
40	2 1078	1 1164	1 0407	2 0264	1 1868	2 0016	2 0356	2 0434
	2 0719	2 1317	2 0861	1 0843	2 0256	1 0517	2 0552	1 1201
	1 0768	2 0977	1 0957	2 0209	2 0891	1 0838	2 0945	2 1318
	2 0689	1 0414	2 0328	2 0518	1 1110	2 2968	2 0558	2 0268
	1 1025	2 0105	2 1194	1 2782	1 1822	2 2026	2 0720	1 0391
45	2 0926	2 1236	2 0014	2 0792	1 0935	2 0600	2 0826	2 0889
	2 0383	2 0265	1 0026	2 0111	2 0416	1 0895	2 0607	1 1241
	2 0607	1 1619	2 0235	1 1913	1 0676	2 1413	1 0111	1 2040
	1 0878	2 1457	2 2911	2 1505	2 0925	2 0101	1 0915	2 0681
	2 0241	1 1116	1 0935	1 0547	1 0489	1 1788	2 0322	2 1459
50	2 0894	2 0476	2 0467	2 1063	1 0265	1 0795	1 3353	2 0569

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 19

row no. col. no.	1	2	3	4	5	6	7	8
	1 1596	1 1231	2 0002	2 0295	2 0747	2 0346	2 0528	2 0775
	1 0512	1 1283	1 0051	2 0429	1 0886	1 2248	1 0292	1 0815
	1 1289	1 0462	1 0192	2 0757	1 0231	1 1692	2 0764	1 1110
	1 0648	2 0800	2 1074	1 0409	2 0436	2 0140	1 0086	1 1807
5	2 1546	1 0244	2 1199	2 0875	2 0424	2 0663	2 0869	2 0061
	2 1847	2 0061	1 0374	2 0195	1 1704	2 1016	1 0200	1 0857
	2 1415	1 0462	1 0013	1 2374	1 0904	2 0375	1 0682	1 0759
	1 0505	2 2066	1 1569	2 0664	1 1516	1 0575	1 1990	1 0409
	1 0320	1 0158	1 0944	1 0011	2 0766	1 0543	2 0955	2 0741
10	2 0424	1 0709	1 0892	1 0186	1 0091	1 1474	1 1500	2 1372
	2 1410	1 0771	2 0672	1 1402	2 0955	2 0393	1 1072	1 0286
	2 0188	2 0803	1 0051	2 0526	2 1560	1 1480	1 0584	1 0568
	1 0204	1 0479	2 1885	1 1016	2 0066	1 1958	2 0249	2 0118
	1 0162	1 0163	2 0389	2 0245	2 0686	2 0630	1 0621	2 0057
15	2 1068	1 2113	1 0651	1 0064	2 0282	1 1183	2 1009	1 1861
	2 0810	1 0195	2 1366	1 0099	2 0322	1 0784	2 0929	2 0251
	1 2462	2 0423	2 0696	2 0746	1 0005	1 0959	1 1117	2 0559
	2 0706	2 0769	2 0128	1 0474	2 1214	1 0162	1 0560	1 1603
	2 1211	1 1194	2 0631	1 1594	1 0661	1 0489	2 1680	2 1466
20	2 0651	1 0154	2 1058	1 0603	2 1638	2 0781	1 0864	2 0397
	1 0294	2 0082	2 0683	2 1390	1 0302	2 0439	2 1655	1 1631
	2 0555	2 1391	1 0253	2 0333	2 0031	1 1571	2 0463	1 1669
	2 0895	1 0092	2 0203	2 2294	1 1790	1 0740	2 0315	2 2319
	2 0112	1 1188	2 0019	2 0081	1 1057	1 0065	1 0605	2 0684
25	2 0111	1 0407	1 0928	1 0645	1 0757	2 1108	2 1547	2 0726
	1 0621	2 0296	2 1881	1 1579	2 0313	1 1744	1 0740	1 0338
	1 1502	2 1629	2 0183	2 0030	2 0462	1 0799	2 0296	1 2084
	1 0085	1 1160	2 1073	1 0027	2 0256	2 0864	1 1192	1 0773
	1 0522	1 0069	1 0964	2 2064	1 0627	2 2115	1 1337	2 0925
30	1 0385	2 0062	2 1468	1 0614	2 0634	2 1003	2 0643	1 0071
	2 0263	2 0431	1 0476	2 0025	1 1404	1 0208	2 0087	1 0469
	1 0845	1 0939	1 1151	1 0504	1 0622	2 0234	2 2079	1 0399
	2 0186	1 0661	2 0191	1 1022	2 1237	2 2706	2 1291	2 0451
	1 0856	2 0755	1 0398	1 0301	1 0252	1 1882	2 0528	2 0104
35	2 2214	1 1362	1 0698	0000	2 0708	1 0822	1 0603	1 0550
	1 1247	1 1561	1 0007	2 2387	2 1650	1 0617	1 0706	2 0166
	1 1591	2 0282	1 0577	1 0161	1 2549	1 0879	2 1365	1 1144
	2 1048	1 0375	2 1840	2 0555	1 0209	2 0285	2 0813	1 0335
	1 1078	1 0634	1 0652	2 0725	1 0766	2 0521	2 0494	1 0444
40	1 0094	2 1224	2 1743	1 1067	2 0231	2 0551	1 0395	1 0925
	1 0355	2 1271	1 2315	1 0575	2 0760	2 0260	2 1478	1 0879
	2 0353	2 0594	2 0308	1 1563	2 2277	1 0890	2 0489	2 0610
	2 0434	1 0824	1 0490	1 0385	1 2144	2 1131	1 0162	2 1125
	2 0067	2 1463	2 0319	1 0289	1 0928	2 0487	2 0152	1 0254
45	1 1351	2 1566	1 1312	2 1545	1 1444	2 1992	1 0087	1 1560
	2 0795	1 0489	2 1025	1 0554	1 0314	2 0681	1 0134	2 1306
	2 1442	2 0782	2 0025	1 0216	1 0938	1 0212	2 0376	2 0514
	1 0572	2 2245	2 1734	1 1070	1 0806	1 0275	2 0513	1 1118
	2 1267	2 0247	2 2495	2 0241	2 0089	2 0415	1 0966	2 0459
50	2 0239	2 1449	1 0217	2 0660	2 0133	1 0239	1 0299	2 0546

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE NO. 20

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 1289	2 0106	1 0211	1 0613	1 1216	1 0218	2 0112	2 2576
	2 1076	1 0339	2 0402	2 0320	2 0203	1 1516	1 0724	2 2217
	1 0729	2 1972	2 1075	1 0522	1 0830	1 0658	2 0831	1 0297
	2 0075	2 0112	1 0396	2 0676	2 0252	1 0669	1 0046	2 0146
5	1 1990	2 0432	2 0159	2 0752	1 1730	2 1878	2 2257	2 0410
	1 0392	1 0358	2 0036	1 0781	2 0439	2 0408	1 1210	2 0826
	2 0744	2 2468	2 0804	2 0942	1 0760	2 0519	2 0830	2 0760
	2 0876	1 0890	1 0023	2 1807	1 1137	2 0677	1 0465	2 0982
	1 0090	2 0886	2 0412	1 0367	1 1075	1 1800	1 1125	1 1171
10	2 1811	2 0588	2 0165	1 0815	2 0869	1 1536	1 0523	1 1241
	2 1896	2 1206	2 0698	1 0571	2 0533	2 0088	2 0642	2 1016
	1 1967	1 1869	1 0262	2 0656	2 0175	1 0406	1 1940	2 1384
	1 0148	2 0596	2 0336	2 0308	2 2442	2 2155	1 1063	1 0158
	1 0806	1 0243	2 1834	2 0898	2 0566	1 0873	1 1430	1 0837
15	2 0604	2 1817	1 0005	2 1847	2 1513	1 0606	1 1147	1 1205
	2 0131	1 1578	2 0792	2 0516	1 0930	2 0522	2 0394	2 0889
	2 0360	1 0843	1 0345	2 1905	2 0743	2 0489	1 1308	1 0959
	1 0076	1 0824	1 0598	1 1413	2 0038	2 1458	1 0451	1 0292
	1 0104	1 0254	2 0077	2 0455	2 0565	1 0242	1 0091	2 0483
20	1 0324	1 0177	2 0096	2 0071	1 0984	2 0354	1 1320	2 1726
	2 0255	1 0132	1 1700	1 0702	2 0545	2 1673	1 0071	2 1323
	1 0399	1 0797	2 0121	2 2073	2 1452	1 0127	1 1235	2 0842
	2 0107	2 0004	1 0409	1 1369	1 0020	2 0848	2 0007	1 1658
	2 0548	1 1026	2 2073	1 0057	1 1264	1 0073	1 0117	2 0919
25	1 1266	1 0284	1 0165	1 0373	1 0458	1 2192	2 0302	2 0084
	2 0758	1 0651	2 0560	2 2099	2 1302	1 0270	2 0718	1 1241
	2 0316	2 0216	1 1003	1 0314	1 0855	2 1627	2 1367	2 0037
	1 0517	2 0838	2 0638	2 0013	1 1572	2 0230	2 0301	1 1282
	1 2130	1 1283	2 1433	2 0297	1 0727	2 0489	1 1560	2 1394
30	2 0453	1 1098	2 1586	2 0270	2 0076	1 0270	1 0049	1 0268
	1 0204	2 0226	2 0972	1 0645	2 0424	1 1927	1 0045	2 0172
	1 0050	1 0210	1 0567	1 1523	1 0509	1 0089	2 1099	1 0765
	2 2257	1 0966	2 1025	2 0075	2 1695	1 0841	2 0473	1 0171
	2 1161	1 0832	2 0909	1 0728	2 0613	1 0683	1 0438	2 1887
35	1 0357	1 0996	1 0055	2 0093	1 0373	1 1192	1 0658	2 1872
	1 0962	2 0031	2 0263	2 1061	2 1302	1 0194	2 0186	1 0125
	1 0684	1 0410	1 0821	1 1793	1 0122	2 0264	2 1064	1 0421
	1 1162	2 0891	2 1639	2 1283	1 0327	2 0349	1 0676	2 0212
	2 0121	2 0866	1 0905	1 0224	2 0362	2 0861	1 0025	1 0461
40	1 0543	2 0658	2 0721	1 0884	1 0803	2 2016	2 0232	1 0791
	1 2294	1 0505	2 1209	2 1528	1 0817	1 0707	2 1206	2 1516
	1 0875	1 0831	2 0804	1 1170	2 1014	1 1009	1 0188	2 0088
	1 1435	1 1435	2 0405	1 0512	2 0711	1 0306	1 0020	2 0686
	2 0725	2 1532	1 0729	2 1788	2 1097	1 0784	2 0608	2 0776
45	2 0626	1 0459	1 1042	1 2054	2 0376	1 0373	1 0919	1 0026
	1 0488	2 2346	1 0722	2 0384	2 0144	1 0365	1 0490	2 1221
	2 1666	2 0188	1 1468	1 0374	2 0158	2 1320	2 0999	2 0795
	1 0882	2 0948	1 0866	2 1239	2 1071	1 0328	2 0443	1 1132
	2 0890	2 1941	1 0677	1 0885	1 0824	2 0754	2 0164	1 1434
50	2 2457	1 2073	2 0541	1 0525	2 0694	1 1394	1 1075	1 0730

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 21

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0003	2 0501	2 0926	2 0553	1 0907	1 0623	1 0746	2 0851
	1 0684	2 1178	2 0090	1 1011	2 0589	1 1558	2 0633	2 0290
	2 0979	1 1071	1 0990	2 0789	2 1887	2 0249	1 0677	1 0077
	2 1073	1 0793	1 0498	1 0658	2 0430	2 0297	2 2007	2 0940
5	2 0009	2 0702	1 1594	2 1745	2 0889	2 1423	2 1890	2 0398
	2 0119	1 1332	2 0723	2 0085	2 0867	1 1532	2 0030	2 0929
	2 0150	2 1184	1 0681	1 0107	1 0508	1 1235	2 0048	1 0702
	2 0622	2 0977	1 0043	1 1091	2 0586	1 0468	2 1861	1 1019
	1 0078	1 0230	2 1301	1 0508	2 0085	2 0319	2 0257	1 0873
10	1 0439	2 0527	1 0558	2 0401	2 0274	2 0348	1 1958	1 0240
	2 0620	2 0797	1 1230	2 0389	1 0961	1 0717	2 0265	2 0687
	1 0833	1 0085	1 0699	2 1550	1 0023	1 0169	2 1689	2 0170
	1 0326	1 0076	1 1803	2 1304	2 0900	2 0192	2 1346	2 1766
	2 1486	2 1539	2 0425	2 0419	1 0894	1 0920	2 1714	1 0280
15	2 0669	2 1154	1 0334	2 1589	1 1332	1 0242	2 1662	1 0674
	2 2543	2 0254	1 0669	2 0972	1 1716	1 0707	1 2137	1 0800
	2 0496	1 0346	2 0523	1 0421	2 1003	2 0344	1 1320	2 0479
	2 0543	2 0519	1 0223	1 0584	2 0582	1 0181	2 1862	1 0510
	1 0309	1 0451	2 1071	2 0418	1 3291	2 0075	1 1176	2 0507
20	2 0662	1 0245	1 0567	1 0225	1 0227	2 0070	1 0920	2 0092
	2 1469	1 0372	2 0272	1 1736	1 0657	1 0136	2 1367	1 3432
	1 1960	1 1100	2 0444	1 0782	1 1120	1 0077	2 2137	1 0432
	2 0191	2 0048	1 0894	1 1238	2 0871	2 0381	1 1489	1 1292
	2 1120	1 0271	1 0138	2 0395	1 1421	1 0833	1 1056	1 0645
25	2 0155	1 1466	2 0352	1 0007	1 0856	1 0611	2 1307	1 0828
	1 0199	1 1315	2 1626	2 1677	1 0125	2 0421	1 1863	2 0454
	1 1945	2 0957	2 0516	1 0847	2 0406	2 0264	2 0583	2 0458
	1 1188	2 0347	1 1591	2 0848	2 0631	2 1309	1 0678	2 0161
	1 0763	1 0152	1 0143	1 2022	1 1033	2 0011	1 2556	1 0263
30	2 0615	1 0819	1 0099	1 0408	2 0532	1 1187	2 0563	2 1029
	1 2257	2 1078	2 0790	2 1016	2 1343	1 0607	2 0349	2 0293
	1 0367	2 0875	1 0282	2 0658	2 0410	2 0193	2 0574	1 0654
	1 0188	1 1041	1 1001	1 0306	2 1706	2 1194	1 0628	2 0960
	1 0268	1 0801	1 0931	2 0685	2 0668	1 0187	1 2032	1 2452
35	2 0010	1 0562	2 0456	2 0302	1 0740	2 0630	1 0011	2 0411
	1 0267	2 0091	1 0554	2 0313	1 0535	1 1979	1 1410	2 0445
	2 1914	2 0264	2 0987	1 0165	1 1203	2 0771	2 0230	2 1977
	2 0127	1 0044	1 0503	2 1315	1 0690	2 0865	2 0521	1 0314
	2 0164	1 0067	2 0821	1 0948	2 0809	1 0177	2 0110	1 0553
40	1 1250	1 0407	2 1469	2 0463	1 0802	2 0296	1 0015	2 0794
	1 0067	1 0214	2 0289	1 0392	1 0546	1 0344	2 0772	2 0254
	1 0592	2 0874	1 0484	1 1048	1 2001	2 1019	1 0524	1 0365
	2 0445	2 0246	2 0231	2 0676	1 0119	1 0102	2 0825	1 0199
	1 0001	2 0299	1 0090	2 1013	2 0277	1 0404	2 1261	1 2353
45	2 0861	1 0045	2 0224	2 0018	1 0947	2 1061	2 1408	2 0694
	2 2229	2 0915	1 0559	1 1460	1 0664	1 1231	2 2071	1 1191
	1 1562	1 0262	1 2968	2 1812	1 0866	2 0839	2 2175	1 2095
	1 0847	1 0515	1 1295	2 1059	2 0630	1 0539	1 0249	2 0843
	1 1379	2 0823	1 0322	1 1471	1 2001	2 0419	2 0072	1 0625
50	2 0611	2 0438	1 1016	2 2706	2 0056	2 0818	2 1070	2 2770

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 22

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0800	1 1039	1 0424	2 0121	1 0923	2 1108	2 1336	1 1495
	1 0771	1 0336	1 1313	1 0622	2 0303	1 1057	1 0886	2 0107
	1 0379	1 1534	1 0773	1 0121	2 0376	1 0820	1 0686	2 0430
	1 1405	2 0027	1 0253	2 0302	2 1138	1 0500	2 1089	1 0035
5	1 0191	1 0219	2 1567	1 0494	2 1661	2 0992	2 1120	2 1067
	1 0391	1 0342	2 2084	2 1191	1 0208	1 0844	1 1486	2 0723
	2 0189	1 0093	1 0285	1 2077	2 1001	2 0457	2 0466	1 1414
	1 1785	1 2183	1 1779	2 0805	1 0443	2 1820	1 0045	1 1037
	2 0238	2 0975	1 0650	2 0238	2 0246	2 0160	2 0193	2 0813
10	1 0239	2 0277	1 2280	2 0347	2 0609	1 1443	1 0487	1 0353
	2 1077	2 1366	2 1334	2 0671	1 1955	1 1370	2 2056	2 1475
	2 0239	2 0131	1 1259	1 1186	2 0998	2 0964	1 2346	1 1418
	2 1287	2 0180	1 0353	2 0821	1 1136	2 0212	2 0518	1 1039
	1 0589	2 0474	1 1211	1 0238	2 0892	2 0584	2 2044	1 0822
15	1 0307	1 0069	1 0016	2 0902	1 0282	2 1383	1 0597	2 0271
	2 0440	1 1042	1 1476	1 1351	1 2395	2 0881	2 0630	1 2054
	1 0872	1 0493	2 1683	1 0109	1 0037	1 0559	2 0445	1 0495
	1 2071	2 0229	1 1416	2 1349	2 1180	2 2241	1 0684	2 0051
	2 0635	2 0287	2 0732	1 1070	2 0326	2 0983	1 1679	2 1501
20	1 0628	2 1890	2 0005	1 1011	2 0369	1 1295	2 1557	2 1688
	1 0610	2 0057	2 1000	2 0126	2 0082	2 0715	2 0733	2 1618
	1 2346	1 2111	2 0735	1 0242	2 0876	2 1189	1 0978	1 0658
	2 1917	2 0463	1 0194	1 0396	1 0591	2 1141	2 0541	1 1488
	1 0492	1 0262	2 0334	1 0227	1 0083	2 1122	1 0422	1 0912
25	2 0247	1 0398	2 0100	1 0058	2 0609	2 1505	1 1333	1 0604
	1 0219	2 1023	1 1002	1 0103	1 1436	2 1368	2 0508	1 0437
	1 0392	1 0989	1 0433	1 0548	2 0127	2 0853	2 0655	1 0624
	1 0774	1 1311	2 0511	2 1672	2 0960	1 0012	1 0274	1 0665
	2 0606	1 0333	1 0110	2 1285	2 0307	1 1013	2 1257	2 0093
30	1 0247	2 0429	2 0307	1 0269	2 0597	1 0036	2 0054	2 0846
	2 0591	1 0810	1 0487	1 1324	2 1818	2 0083	1 0458	1 0867
	1 0255	2 2597	1 0038	2 0060	2 0648	1 0412	1 0411	2 0227
	2 0379	1 0475	1 0269	1 0057	2 1266	1 0010	2 0144	1 1226
	2 0179	1 0584	2 0122	1 0213	1 0488	1 1058	1 0479	1 0929
35	2 0139	2 0489	2 1263	2 0828	1 1731	1 1170	1 0651	2 0275
	2 0724	2 0188	1 1986	1 1812	1 1004	1 0407	1 0054	2 1206
	1 1596	1 1198	1 1308	1 1482	1 1654	2 0667	2 2387	2 1059
	1 2414	2 0625	1 0532	2 0933	1 0637	1 0550	2 0275	2 0857
	1 0128	2 0898	2 0239	1 0917	2 0385	1 1779	2 0717	2 0928
40	1 0958	1 0561	1 0980	1 0703	1 0052	1 0049	1 0364	1 0552
	1 1544	1 0971	1 0951	1 0179	2 0222	1 0133	1 1208	2 1634
	1 0149	2 0981	2 0053	2 1281	2 0636	1 0114	1 0835	1 1529
	1 0218	2 0013	2 0318	1 0505	2 0750	1 0580	2 1114	1 0380
	2 0222	1 0087	2 0831	2 1906	2 0434	2 0997	1 0825	2 0614
45	2 1752	1 2040	1 0283	2 0810	1 0016	1 1020	1 1141	1 0215
	2 0881	2 0365	2 1158	2 0859	2 0451	1 0253	2 2404	2 1431
	2 0263	2 1171	2 0062	2 0777	2 1891	1 1138	2 0975	2 1158
	1 0822	1 0128	1 1158	2 0348	2 1928	1 1104	2 0520	1 0184
	1 0325	2 0030	1 0652	1 1279	1 1281	2 0586	1 0024	2 0716
50	2 1371	1 2144	1 0140	2 0005	2 0177	1 1821	2 0516	2 0706

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '—' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 23

row no. \ col. no.	1	2	3	4	5	6	7	8
	1 0633	2 0280	2 0289	2 1340	1 1030	2 1366	1 0990	1 0436
	2 0534	1 2284	2 0129	2 0110	1 0213	1 0204	1 0013	2 0621
	2 0119	1 1421	1 0945	2 0821	1 0035	1 1076	1 1150	2 0408
	1 1410	1 0112	1 0326	1 0444	2 0846	1 0938	1 0333	2 0082
5	1 0054	2 1782	2 0774	1 0429	1 1161	1 1314	2 0204	2 0772
	1 0310	1 0508	2 0871	2 1353	1 1114	2 0038	2 1460	2 0429
	2 2989	1 0220	2 1587	2 0257	2 0490	1 0150	1 0043	2 1211
	1 0300	1 1423	2 0738	2 1398	1 1383	2 1379	1 0630	2 0140
	2 0418	2 0839	2 0454	1 1789	2 0384	1 0878	1 1189	2 1154
10	2 0218	1 2214	1 0247	1 2081	2 1288	2 1255	2 0819	1 0734
	2 0472	1 2189	1 0075	1 0791	1 0480	1 1577	2 1238	1 0402
	2 0114	1 0584	2 1073	2 0632	2 0669	1 0601	1 0916	1 1363
	1 1460	2 0881	2 0442	2 0529	2 0441	2 0334	1 0178	2 0175
	2 0673	1 0005	2 0452	2 0454	1 0607	2 1894	1 0393	1 0051
15	2 0197	1 1434	2 0769	2 0506	1 0240	2 0510	2 1762	2 1605
	1 0122	1 0347	2 0658	1 1407	2 0142	1 0847	1 0079	1 0543
	2 0803	2 2878	1 2183	2 1385	1 1335	2 0732	2 0422	2 0139
	2 0320	2 0207	2 0604	1 0077	2 0152	2 2217	1 0908	2 0808
	2 1584	2 1821	2 0746	1 2178	2 1960	2 0973	1 1428	1 0144
20	1 0937	1 1411	1 0823	1 1214	2 0929	1 1137	1 0319	1 1027
	1 0347	2 0188	2 0243	2 0255	2 1879	2 0087	2 0956	1 1465
	2 0475	1 1970	2 0563	1 0652	1 0864	2 1678	1 0064	1 1101
	2 0145	1 0783	2 0156	2 0397	1 0304	2 0030	2 1409	2 0712
	2 0342	1 0207	1 0429	1 0229	2 1004	2 1031	2 1006	1 0302
25	1 0296	2 1040	1 0431	2 1962	1 0248	2 1266	1 1694	2 0493
	2 0863	2 0614	1 0061	2 0084	2 0298	2 1849	1 0103	1 0516
	2 0636	2 1614	2 1504	2 0256	2 0566	2 0371	1 1614	2 0516
	1 0308	2 1202	1 0756	2 0807	1 0064	1 0492	2 0005	1 1281
	1 0756	1 1941	1 1352	2 0850	2 1725	1 1203	2 1861	1 0508
30	2 1096	2 0258	2 2209	2 0655	1 1824	2 0254	1 1719	2 0084
	2 0282	2 0476	2 1376	2 0803	2 0177	1 0383	1 0747	2 0510
	1 0784	1 0534	2 0359	2 2457	2 1073	2 0036	2 1706	1 0396
	1 1034	1 0244	1 0340	2 1262	2 0127	1 0206	2 2506	2 1875
	2 0455	1 1204	1 1815	2 0448	1 0419	1 0209	1 0189	1 2075
35	1 0684	2 1126	1 0252	2 0611	1 0795	2 1501	2 0600	1 0961
	2 0494	2 1143	1 0154	1 1213	2 0098	2 0238	1 0360	2 1594
	2 0278	2 0865	1 0048	1 0281	2 0047	1 0447	1 0394	1 0257
	1 0367	2 1566	2 0206	2 0434	2 2400	1 0006	1 0716	1 0609
	2 2106	2 1195	2 1291	1 0055	1 0965	2 0340	1 0002	1 1454
40	1 1276	2 0475	1 1631	2 0880	1 0273	1 0531	1 0496	1 0910
	1 0737	1 1079	2 0135	1 0494	2 0276	2 0882	2 0953	1 0448
	1 0389	2 0638	1 0996	1 0410	2 0304	2 1178	1 2162	1 1059
	1 0462	2 0513	1 1652	2 0143	1 0008	2 0922	2 1105	1 0628
	2 1024	2 0297	2 0305	2 0809	1 0055	1 2206	1 1544	1 1784
45	2 0597	1 0868	1 0752	2 1423	2 2223	1 0565	2 1378	2 0713
	2 1032	2 0070	2 0429	1 0524	1 0385	1 0185	1 1346	1 1305
	2 0233	2 1477	2 0885	2 2226	1 0838	1 0640	2 0670	2 0582
	2 0010	1 1119	1 0202	1 1530	2 0384	1 1283	2 0938	2 0307
	2 0100	1 0525	1 0445	2 0644	2 0328	2 0142	2 0407	1 0635
50	2 0276	2 2669	1 0063	2 0022	2 0215	1 0646	1 0864	2 0229

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 24

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 1957	1 0382	2 0868	1 0149	2 0409	2 0244	1 2197	1 0113
	1 0125	2 0344	2 0818	2 0564	2 1486	1 2037	2 0571	1 0962
	1 0043	2 0864	1 1735	1 0823	2 0500	1 0316	1 0476	1 1349
	2 1840	2 0249	1 0883	1 1306	2 0874	1 0009	1 0487	2 0306
5	2 0805	2 0388	2 0532	1 0070	1 1347	2 1643	2 2084	2 0752
	1 0361	2 0773	2 1772	1 0901	1 0835	2 0954	1 0624	2 1653
	2 0058	2 0781	1 1997	2 1745	2 1181	2 0476	2 0158	1 0711
	2 0474	1 1157	1 1169	1 0567	1 0692	2 1237	1 0653	1 1044
	2 0206	2 1417	2 2576	1 0018	2 1071	2 1770	2 1680	1 1282
10	2 0590	2 1022	2 0651	1 0557	2 0316	2 0637	2 0827	2 1256
	1 0081	1 1902	1 0743	1 1023	2 0535	2 1475	2 0875	1 0476
	2 0085	2 0104	1 0112	2 0304	1 0084	1 0293	1 0104	2 0565
	1 0443	2 0554	1 0756	1 0456	2 0196	1 1413	1 1695	1 0114
	1 0239	1 0750	2 0998	1 1738	1 1040	1 0607	1 0607	2 0575
15	2 1026	1 0331	1 0726	1 0185	2 0021	1 0234	2 0912	1 1480
	1 2251	2 0730	1 0444	1 0031	1 0497	2 0757	2 0600	2 0330
	2 0201	2 0850	1 0985	1 0699	2 0600	2 0919	2 0798	2 2512
	1 0320	1 0545	1 0636	1 0565	1 0005	1 1344	2 0070	2 0539
	2 0281	1 1356	1 1335	1 0715	1 0425	2 0274	1 0149	1 1927
20	2 0952	1 0191	2 0191	1 1977	1 1203	1 0184	1 0969	2 0795
	1 1848	2 0088	2 0184	1 0862	2 0516	1 0030	1 0408	1 0437
	1 0607	1 0562	2 1184	2 0207	1 0938	2 0337	1 0307	2 1010
	1 1631	2 2034	2 1349	2 0402	2 0203	1 1632	2 1731	2 1928
	2 0924	1 0795	1 0478	2 1362	2 0039	2 1945	2 0008	1 1125
25	2 1175	2 1787	1 0380	1 1760	1 1209	1 1618	2 0066	2 0649
	2 1061	2 0070	2 0150	1 0175	2 1259	1 0130	1 2048	2 0017
	2 1789	1 0286	1 3121	1 0086	1 0933	1 0588	1 0584	2 0156
	1 1592	1 0039	2 1986	1 1745	2 0087	1 0792	2 0183	1 0269
	2 0555	2 1137	1 1211	2 1913	1 0036	2 0947	2 2209	2 1193
30	1 0007	2 1684	1 0108	1 0759	1 2084	2 0145	1 1131	2 0730
	2 0322	1 0003	2 0428	2 0191	1 0827	2 0080	1 0500	1 0991
	1 1181	2 1412	2 0918	1 2290	2 1134	1 0300	2 0233	2 1120
	2 0543	1 0124	1 0403	1 0198	2 1680	2 1100	2 0248	1 0174
	2 0684	2 0094	2 1055	2 0542	1 0062	1 1008	2 0712	1 1656
35	2 0257	2 1571	2 0191	0000	2 0641	2 1967	1 1741	1 0801
	2 1317	1 0544	2 1098	2 0647	2 0456	1 1074	1 0693	2 0202
	2 0356	2 0254	2 0092	2 1696	1 0453	1 0833	2 0870	2 0134
	1 0602	2 0578	2 0578	1 0011	1 0736	1 1102	1 0487	1 0733
	1 0480	1 0630	2 0228	2 1204	1 1229	2 1525	2 0125	1 1509
40	2 0594	2 0939	2 0184	2 1019	2 0054	1 0937	1 2052	1 0249
	2 0148	2 0594	2 0555	1 0927	1 0782	2 0080	2 0435	2 1314
	1 1496	1 1330	2 1970	2 1866	2 0584	1 0486	1 2620	1 1460
	1 0745	2 0789	2 1372	1 0641	2 0778	2 0735	1 1385	1 0831
	1 0528	1 1584	2 0777	1 0017	2 0198	2 0685	1 0489	1 2030
45	2 0262	2 0527	1 0148	1 1896	2 0502	2 2661	2 0820	2 0468
	1 0371	1 1009	2 0390	1 1390	1 0609	1 1085	2 1141	2 1322
	2 0073	2 0403	2 0014	1 0152	1 0203	1 0828	2 0393	2 1403
	2 0627	2 1497	2 1427	1 1354	2 1211	2 0717	2 0148	2 0382
	1 0248	2 0609	2 0401	2 0104	1 0248	1 0538	1 0649	2 0542
50	1 1486	2 1576	2 0225	2 1627	1 1156	2 0348	1 1030	2 1492

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '−' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

TABLES OF RANDOM NORMAL DEVIATES

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

PLATE No. 25

row no. \ col. no.	1	2	3	4	5	6	7	8
	2 0055	2 0349	1 1566	2 0999	1 0065	2 0763	1 0600	1 0694
	1 1123	1 0134	1 0192	1 1353	2 0292	1 0113	2 0756	2 1183
	1 0414	1 0532	1 0139	2 0525	2 0619	2 0732	2 1826	2 0103
	2 0466	1 1167	1 0877	2 1784	1 0786	2 1906	2 0122	1 1394
5	2 1017	2 0476	1 0606	1 0199	2 0409	2 1404	1 1869	1 0138
	1 0674	2 1376	2 0133	2 1042	1 0173	1 0904	2 0438	1 1359
	1 1176	2 1900	2 0210	1 0733	2 1144	2 0071	2 1722	1 1721
	1 0465	1 0365	1 0035	2 1655	2 0301	1 0953	1 0254	2 1369
	1 1676	2 0752	1 1147	1 0456	1 0623	2 0925	2 0149	1 1356
10	2 1668	1 0129	1 1063	2 1288	1 2447	1 0529	2 1039	1 1132
	2 0223	1 0402	1 0305	2 1121	1 1254	2 0107	2 1881	2 0487
	1 0305	1 1690	2 0861	1 1554	2 0374	1 0653	1 1245	2 0159
	2 0655	2 0878	1 0085	2 0336	2 1201	2 0394	1 1122	1 0910
	2 1133	2 1224	1 0325	2 0174	2 1721	1 0668	1 0824	1 1106
15	2 0114	2 0308	1 0734	1 0121	2 0496	2 0047	2 0480	2 0517
	1 1478	1 0214	1 1186	1 0137	2 0380	2 1430	2 0187	2 1185
	2 0927	2 0274	2 0439	1 1338	1 0448	1 1063	1 0768	1 0431
	2 0506	1 0174	1 0370	2 0550	2 0154	2 0479	2 0057	1 0868
	2 0486	2 1894	2 0620	1 1198	1 1231	2 0499	1 0951	1 1792
20	1 0618	2 2267	1 1048	1 0202	2 0074	1 0297	2 1002	2 0661
	1 0080	1 0436	2 2549	1 0148	1 0203	2 1184	2 0714	1 0587
	2 0355	2 1059	2 1326	2 0444	1 0225	2 0220	2 0208	1 0356
	2 0549	1 1218	1 0275	1 0729	1 0174	1 0354	2 0405	2 0014
	2 0591	2 0668	1 0496	1 1249	1 3011	1 0671	2 0285	1 0338
25	1 0352	1 0211	1 0170	2 1753	1 0274	2 0360	1 0874	2 1055
	2 0135	2 0612	2 0798	1 3719	2 0780	2 0985	2 0632	2 1546
	2 0266	2 0803	1 1626	2 0188	2 0174	2 0375	1 1519	1 0748
	1 1155	1 0025	1 0203	1 1089	1 1076	1 0352	1 0264	1 0525
	2 0693	1 1184	2 0408	2 0471	1 0069	2 2036	1 1469	1 0054
30	2 1911	1 0268	2 1110	2 0148	2 2095	1 1917	1 0240	2 1083
	1 0223	2 0585	2 0600	2 0050	1 1236	1 0026	2 0093	1 0667
	1 0937	2 1031	2 0252	1 0047	2 0027	2 0143	2 1718	2 1289
	1 0990	2 3353	2 0700	1 1498	1 1815	2 0369	1 0208	1 1960
	1 1240	2 0537	2 1134	2 0031	2 0010	1 0091	1 0018	1 2040
35	1 0566	1 1649	1 0864	1 0847	2 1890	1 0279	2 0168	2 0966
	1 0199	2 2644	1 0250	1 0814	1 1070	1 0418	2 0076	1 0700
	2 0084	1 0646	2 0674	2 0064	1 1149	1 0283	1 0291	1 0996
	2 1246	2 0317	2 1167	2 1606	1 0648	2 0802	1 0534	2 0585
	2 1754	2 2104	1 0926	1 0496	2 0741	1 0460	2 0228	1 0487
40	2 0780	2 0938	2 0663	2 0623	1 0877	1 0390	1 1008	2 2326
	2 1435	2 0313	1 1571	1 1852	2 1161	2 2597	2 0648	1 1272
	1 1183	1 0672	2 1957	1 0086	2 0042	1 0387	2 0985	2 0073
	1 0031	1 2257	2 0010	2 0330	2 0307	1 0675	1 1020	2 1809
	2 2077	2 0804	1 1246	2 1138	2 0256	2 0751	1 0139	1 0533
45	1 0219	1 0024	1 0332	1 1027	1 1031	2 0873	1 0454	2 0109
	1 0803	1 1726	1 1052	2 0241	1 0093	1 0327	1 2462	1 0060
	2 0616	2 0462	2 0352	2 1460	2 0517	1 1309	2 0176	1 0327
	2 1041	1 0764	1 0634	1 0230	2 0573	2 0582	1 0957	2 1527
	2 1599	2 0625	1 1092	1 0281	2 0017	1 0748	1 0572	1 1760
50	2 0376	1 0464	1 0013	1 0366	2 0599	1 0991	1 1935	2 1019

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

RANDOM NORMAL DEVIATES BASED ON TIPPETT'S RANDOM SAMPLING NUMBERS

Plate No. 26

col. no. row no.	1	2	3	4	5	6	7	8
	2 2162	1 0539	2 0748	2 1099	2 0432	2 2989	1 0460	2 1137
	1 0327	2 0913	1 0009	1 0446	2 0296	1 0518	1 0549	2 0255
	1 0688	2 0665	2 1865	2 0530	2 2346	2 0629	1 0954	1 0041
	1 0758	1 0108	2 0034	1 0385	2 0218	1 0894	2 1890	2 0552
5	2 0214	2 0285	2 0507	1 0708	1 0892	2 0107	2 0757	1 0334
	2 1430	1 1520	2 0633	1 1083	2 0819	2 1449	2 0511	2 1545
	2 0078	2 0344	1 1845	2 1902	1 0475	1 0780	1 0196	2 0619
	2 0239	2 1431	1 0377	2 0602	1 0021	2 0292	1 0661	1 0184
	1 0122	2 1426	1 0589	2 0022	2 2501	2 1547	1 1469	2 0239
10	2 2241	1 1741	2 0324	1 0792	1 0038	1 0583	2 1123	1 1797
	1 0148	2 0613	2 0598	2 1687	2 2251	2 0070	2 0263	2 0549
	1 0828	2 0864	1 0254	2 0128	1 0124	2 0578	1 0473	1 0559
	1 1223	2 1602	1 0460	1 0995	2 1099	1 2484	1 1078	1 0679
	1 1335	2 0532	2 1321	2 0277	1 0273	1 0635	2 1286	1 1254
15	1 0837	2 2220	1 1118	1 1066	2 0715	2 0198	2 0365	2 0208
	2 0620	1 0285	2 0163	2 0469	1 1039	2 0644	1 0392	2 0772
	2 0944	1 0091	1 0080	1 1762	2 0112	2 1433	1 0932	1 0645
	2 0073	1 0804	2 0452	2 0673	2 1703	1 0857	2 0949	2 1565
	1 0724	1 0207	2 1068	2 0971	1 0569	1 0622	2 0194	1 0772
20	2 1037	2 0551	1 0304	1 0958	2 0076	1 0144	1 0703	1 1529
	2 3036	1 1045	1 1680	2 0300	2 2512	2 0137	2 1475	1 0449
	2 0223	1 0467	2 0144	1 0501	1 0532	2 0630	2 0362	1 1876
	2 0236	2 1999	1 0309	2 0496	2 0016	2 0907	1 0492	1 1486
	2 1571	1 0209	2 0768	2 0623	2 0576	2 1430	2 1455	2 1338
25	2 1013	1 2366	2 0232	2 2678	1 1868	2 1986	2 0379	1 0054
	1 0638	1 0310	2 0402	1 1337	1 0190	1 1671	1 0606	2 0902
	2 0054	2 0229	1 0163	1 1223	1 0004	1 1152	2 1356	2 0118
	1 0797	1 0057	2 0456	1 1014	2 0434	2 0097	2 0120	2 1567
	2 0740	2 0500	1 1264	1 0873	2 1066	2 0510	1 1081	2 0225
30	2 1367	1 1373	1 1099	1 1849	2 1699	2 1112	2 1134	2 0017
	1 3353	2 0548	1 0276	2 1419	1 0977	2 0669	1 0238	1 0805
	2 0075	2 0663	1 0096	2 1100	1 0510	2 1042	2 1288	2 1788
	1 0616	2 0225	1 0004	2 0350	1 0142	1 1039	2 0260	2 1668
	2 2948	1 0353	2 0340	1 0191	2 0033	1 2807	1 1128	2 1050
35	2 0301	2 0099	2 0254	1 1052	2 0385	2 1536	2 0679	2 0645
	2 1442	1 0114	1 1447	2 1073	2 0782	2 1621	1 0257	1 1580
	2 0575	1 0489	2 0653	2 0241	2 0816	2 1165	2 1561	2 0078
	2 0535	2 0463	1 0836	1 1107	2 0593	2 0301	1 0248	1 1380
	2 1011	2 0523	2 1351	2 0726	2 1097	2 0197	2 2077	2 0138
40	2 1401	2 0852	2 0220	1 0841	2 0940	1 0061	2 0199	2 1537
	2 0594	1 1829	2 0315	2 0749	2 1251	1 0703	1 1539	2 0754
	1 0216	2 1162	1 1034	1 0589	2 0580	1 1730	2 0200	2 1107
	2 2442	1 1009	1 0955	2 0277	2 0610	2 1790	1 1383	1 0125
	2 1858	1 0009	1 0446	1 1696	2 0305	2 0830	2 0029	2 0718
45	2 0994	2 1818	1 1034	1 0481	2 0117	1 0137	1 1690	1 1011
	2 0372	2 0342	2 1677	2 0289	1 0001	1 0833	2 0239	2 0780
	2 0714	1 0047	2 0615	2 1495	1 0380	1 1435	2 1736	1 1110
	1 0768	1 0379	2 1515	1 0902	1 0710	2 0722	1 0154	2 2088
	2 0048	1 0978	1 1149	1 0373	1 1794	1 0073	2 0886	2 1341
50	2 0364	1 1471	1 0429	2 0120	2 2162	1 0997	2 1262	1 2319

The figure "1" at the beginning of each group of five figures stands for the '+' sign and the figure "2", for the '–' sign. The normal deviates listed as above have a standard deviation of 1000. To obtain deviates with unit s.d. a decimal point has to be placed after the first digit.

MISCELLANEOUS

ON SAMPLING WITH AND WITHOUT REPLACEMENT

By D. BASU

Indian Statistical Institute, Calcutta

SUMMARY. Certain aspects of sampling with or without replacement, with equal or unequal probabilities, are considered here in some details. Some comparisons have been made between the with and without replacement sampling schemes. When we are sampling with replacement the estimate should not depend on the number of times that any particular unit may appear in the sample. Thus, certain estimation procedures in current use are shown to be inefficient.

1. Introduction

Suppose a given population has N units. Let Y_j be some real valued characteristic of the j-th population unit ($j = 1, 2, ..., N$). Consider the problem of estimating the population mean

$$\bar{Y} = N^{-1} \Sigma\ Y_j.$$

Let

$$\sigma^2 = N^{-1}\ \Sigma\ (Y_j - \bar{Y})^2$$

be the population variance.

If we draw a sample of size n from the population with equal probabilities and with replacement then the variance of the sample mean is $n^{-1}\sigma^2$. If, on the other hand, we draw a sample of the same size n but this time without replacement then the variance of the sample mean is $n^{-1}\sigma^2(N-n)(N-1)^{-1}$. Thus, it is usually claimed that sampling without replacement is better as it leads to an estimator of $\bar{Y}$ with a smaller variance. A little reflection, however, will show that this comparison between the two methods of sampling is not usually quite fair. Let us take a simple example. Suppose the units are villages and Y_j the number of households in the j-th village. Here the cost of selecting the sample villages from a given frame is negligible compared to the cost of travelling to the selected villages and ascertaining the exact number of households in the selected villages. Generally speaking, we need only consider the cost of measuring the Y-characteristics of the selected units—the small cost involved in selecting the units from a frame may be taken to be a part of the large overhead cost. In sampling with replacement it is then the number ν of distinct units appearing in the sample (and not the sample size n) that roughly determines the cost.

2. The distribution of ν

If ν be the number of distinct units appearing in a sample of size n drawn with equal probabilities and with replacements from a population with N units then it is clear that the distribution of ν depends only on n and N. It is not difficult to show (Feller p. 92) that

$$P(\nu = s) = N^{-n} \tbinom{N}{s}[1-\tbinom{s}{1}(s-1)^n+\tbinom{s}{2}(s-2)^n\ \ldots]$$

where s runs from 1 to the smaller of n and N.

In terms of the 'differences of zeros' we may write the above in the more elegant form

$$P(\nu = s) = N^{-n} \binom{N}{s} \Delta^s 0^n \qquad \dots \quad (2.1)$$

where Δ is the usual difference operator with unit increments and $\Delta^s\, 0^n$ is to be interpreted as $\Delta^s\, x^n$ at $x = 0$.

From (2.1) we have the probability generating function of ν as

$$P_\nu(t) = E\, t^\nu = N^{-n} \sum_{s=0}^{N} \left[\binom{N}{s} t^s \Delta^s 0^n\right] = N^{-n} (1+t\Delta)^N\, 0^n. \qquad \dots \quad (2.2)$$

(Note that $\Delta^s\, 0^n = 0$ for $s = 0$ and $s > n$).

Writing $1+t$ for t in the probability generating function we have the factorial moment generating function of ν as

$$F_\nu(t) = N^{-n} (\mathcal{E}+t\Delta)^N 0^n \qquad \dots \quad (2.3)$$

where $\mathcal{E} = 1+\Delta =$ the usual increment operator with unit increments.

$$\therefore \quad E(\nu) = N^{-n} \binom{N}{1} \mathcal{E}^{N-1}\, \Delta\, 0^n = N^{-n} \binom{N}{1} (\mathcal{E}^N - \mathcal{E}^{N-1})\, 0^n = N \left[\, 1 - \left(\frac{N-1}{N} \right)^n \right]. \qquad \dots \quad (2.4)$$

Also
$$E\nu(\nu-1) = N^{-n} \binom{N}{2} 2\mathcal{E}^{N-2}\, \Delta^2\, 0^n = N^{-n}\, N(N-1)(\mathcal{E}^N - 2\mathcal{E}^{N-1} + \mathcal{E}^{N-2})\, 0^n$$
$$= N(N-1) \left[1 - 2 \left(\frac{N-1}{N} \right)^n + \left(\frac{N-2}{N} \right)^n \right]$$

$$\therefore \quad V(\nu) = N \left(\frac{N-1}{N} \right)^n - N^2 \left(\frac{N-1}{N} \right)^{2n} + N(N-1) \left(\frac{N-2}{N} \right)^n. \qquad \dots \quad (2.5)$$

3. SAMPLING COST CONSIDERATIONS

If we assume that the variable part of the cost of sampling is proportional to the number of distinct units in the sample then we may compare the two methods of estimating $\bar{Y}$, as described in § 1, in the following manner. The expected sampling cost for a sample of size n with replacement is equal to the sampling cost for a sample of size $E\nu = N \left[\, 1 - \left(\frac{N-1}{N} \right)^n \right]$ without replacement (let us conveniently forget the fact that $E\nu$ is not necessarily an integer). The variances for the sample means for the two cases are then $n^{-1}\sigma^2$ and $(E\nu)^{-1}\sigma^2(N-E\nu)(N-1)^{-1}$ respectively. A little computation will show that the former is larger. Thus, from this comparison, sampling with replacement appears to be worse than sampling without replacement. This comparison between the two methods heavily depends on the assumption of linearity of the cost function and as such is not very satisfactory. For a different cost function sampling with replacement may appear to fare better than sampling without replacement. The issue that is raised in the next section is perhaps more pertinent to the problem.

4. Two estimators from a with replacement sample

If we draw a sample of size n with replacements and with equal probabilities and if ν be the number of distinct units appearing in the sample then the average Y-characteristic of the ν distinct units is also an unbiased estimator of $\bar{Y}$. We may enquire whether this estimator is better or worse than the average over all the n units. The variance of the former is

$$E\left(\frac{N-\nu}{N-1}\cdot\frac{\sigma^2}{\nu}\right) \qquad \ldots \quad (4.1)$$

whereas that of the latter is $n^{-1}\sigma^2$.

It is not possible to give a simple expression for (4.1). In the next section we shall give an indirect proof of the inequality.

$$E\left(\frac{N-\nu}{N-1}\cdot\frac{\sigma^2}{\nu}\right) < \frac{\sigma^2}{n} \quad (\text{if } n > 1) \qquad \ldots \quad (4.2)$$

Thus, the average Y-characteristics of the ν distinct units in the sample is a better estimator than the average over all the n units. This result may at first appear to be a little unfamiliar, even surprising. Let us take, for example, the familiar Binomial model where we draw n balls at random one by one and with replacements from an urn containing N identical balls Np of which are white, the rest being black. Here the sample observation consists of a sequence of n white or black balls. The number r of white balls in the sample then constitutes a complete[1] sufficient statistic for the unknown parameter p. (Note that in this situation the distribution of the sample depends only on p and not on N.) Hence r/n is the uniformly minimum variance unbiased estimator of p. Now suppose that the N balls are distinguishable from one another (as for example when the balls are villages) or suppose we put distinguishing marks on the balls drawn before they are replaced. The sample observation then is a sequence of n balls and there are N^n possible sample observations each having the same probability. (In the previous case the probability of getting a particular sample observation was $p^r(1-p)^{n-r}$ where r is the number of white balls in the sample). Here the sample observation is more detailed than in the previous case and actually contains more information about the parameter p. Now r is no longer a sufficient statistic. The ν distinct balls that came in the sample is a sufficient statistic and nothing less than this can be sufficient. Consider now a third kind of sample observation where for each of the n balls that are drawn we note down only its colour and the fact whether this particular ball was drawn before or not. Here the sample can be represented as a sequence of n whites and blacks with cross marks at ν places (ν a variable) to indicate at which draws we had the distinct balls. The sample observation now is more detailed than the first case and less so than the second. If ρ be the number of distinct white balls then the statistic (ρ, ν) is sufficient. The conditional expectation of r/n for fixed values of (ρ, ν) is ρ/ν and so by the Rao-Blackwell theorem ρ/ν is better than r/n. Here the statistic (ρ, ν) though sufficient is not complete. This is obvious from the fact that the distribution of ν is independent of the parameter p. Thus we are unable[1] to prove that ρ/ν is the best unbiased estimator of p. The proof sketched above only demonstrators that, with the additional information of which are the distinct units, the standard estimator r/n is no longer the best estimator and that it is in fact worse than ρ/ν. In the next section we give a proof of inequality (4.2) in the general case.

---

[1] The distribution of r is complete if there are at least $n+1$ admissible values for p. Thus if N be smaller than n then the distribution of r will not be complete.

5. PROOF OF (4.2.)

Let there be N population units and let Y_j be the Y-characteristic of the j-th population unit ($j = 1, 2, \ldots, N$). A sample **S** of size n is drawn one by one, with equal probabilities and with replacements. Let y_i be the observed Y-characteristic of the i-th sample unit ($i = 1, 2, \ldots, n$). For each sample unit suppose we also note down its unit index (if a particular sample unit happens to be the j-th population unit then its unit-index is j). Let u_i be the unit-index of the i-th sample unit and let $\mathbf{x}_i = (y_i, u_i)$. We can then record the sample observation as

$$\mathbf{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$$

where the $\mathbf{x}_i'$s are independently and identically distributed random vectors.

Let ν be the number of distinct sample units and let $u_{(1)} < u_{(2)} < \ldots < u_{(\nu)}$ be their unit-indices written in an ascending order. Let $y_{(i)}$ be the Y-characteristic of the sample unit with unit-index $u_{(i)}$ and let $\mathbf{x}_{(i)} = (y_{(i)}, u_{(i)})$ $i = 1, 2, \ldots, \nu$.

Consider now the set of 'order-statistics'

$$\mathbf{T} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)} \ldots, \mathbf{x}_{(\nu)}).$$

The usual estimator of $\bar{Y}$ (the population mean) is based on all the n observations and is

$$\bar{y} = \bar{y}(\mathbf{S}) = n^{-1} \sum_{1}^{n} y_i$$

whereas the estimator based on the ν distinct units is

$$\bar{y}^* = \bar{y}^* (\mathbf{T}) = \nu^{-1} \sum_{1}^{\nu} y_{(i)}.$$

Now, for fixed **T**, the conditional distribution of $\mathbf{x}_i$ is concentrated at the ν points $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(\nu)}$ with equal probabilities at each of these points.

$$\therefore \quad E(y_i | \mathbf{T}) = \bar{y}^* \quad (i = 1, 2, \ldots, n)$$

and hence

$$E(\bar{y} | \mathbf{T}) = \bar{y}^*.$$

Since $\bar{y}$ is an unbiased estimator of $\bar{Y}$, it follows at once that $\bar{y}^*$ is also unbiased. It also follows that, for any convex (downwards) loss function, $\bar{y}^*$ has a uniformly better risk function than $\bar{y}$. In particular

$$V(\bar{y}^*) \leqslant V(\bar{y})$$

the sign of equality holding only when $n = 1$.

Thus the inequality (4.2) is proved. We may note in passing that **T** is a sufficient statistic here though not a complete one. No uniformly best unbiased estimator of $\bar{Y}$ exists.

---

[1] If the parameter N is also unknown then it is easily demonstrated that the distribution of ν is complete. In this situation we believe that ρ/ν is the best unbiased estimator of p.

6. THE CASE WHEN ν IS FIXED IN ADVANCE

In the previous sections we fixed n and had ν as a random variable. Here we consider the situtation where the number ν of distinct units in the sample is fixed in advance. We go on drawing samples one by one, with equal probabilities, and with replacements until we get n distinct units. The probability distribution of the number n of samples drawn may be obtained as follows. The event $n = k$ means that in the first $k-1$ draws there are exactly $\nu-1$ distinct units and that the k-th unit drawn is different from the $\nu-1$ distinct units that appeared in the first $k-1$ cases. Thus from (2.1) it follows that

$$P(n = k) = [N^{-(k-1)} \binom{N}{\nu-1} \Delta^{\nu-1} 0^{k-1}] \left(1-\frac{\nu-1}{N} \right) = \binom{N-1}{\nu-1} \Delta^{\nu-1} \left(\frac{x}{N} \right)^{k-1} \Bigg|_{x=0} \quad \dots \quad (6.1)$$

where $$k = \nu, \nu+1, \dots \text{ ad inf.}$$

From (6.1) it follows that the probability generating function of n is

$$P_n(t) = E\, t^n = \sum_{k=\nu}^{\infty} t^k P(n = k) = \binom{N-1}{\nu-1} t \sum_{k=\nu}^{\infty} \Delta^{\nu-1} \left(\frac{xt}{N} \right)^{k-1} \Bigg|_{x=0}$$

where Δ operates on x.

Since $\Delta^{\nu-1} x^r = 0$ for $r < \nu-1$ we have

$$P_n(t) = \binom{N-1}{\nu-1} t\Delta^{\nu-1} \left[\sum_{k=1}^{\infty} \left(\frac{xt}{N} \right)^{k-1} \right]_{x=0} = \binom{N-1}{\nu-1} t\Delta^{\nu-1} \left(1-\frac{xt}{N} \right)^{-1} \Bigg|_{x=0} . \quad \dots \quad (6.2)$$

In a like manner we have

$$E(n) = \binom{N-1}{\nu-1} \Delta^{\nu-1} \left(1-\frac{x}{N} \right)^{-2} \Bigg|_{x=0} . \quad \dots \quad (6.3)$$

If $\bar{y}$ be the average Y-characteristic of all the n observations then $\bar{y}$ is an unbiased estimator of $\bar{Y}$ with variance

$$\begin{aligned} V(\bar{y}) &= E\frac{\sigma^2}{n} = \sigma^2 \sum_{k=1}^{\infty} \frac{1}{k} P(n = k) \\ &= \sigma^2 \binom{N-1}{\nu-1} \Delta^{\nu-1} \sum_{k=1}^{\infty} \frac{1}{k} \left(\frac{x}{N} \right)^{k-1} \Bigg|_{x=0} \\ &= \sigma^2 \binom{N-1}{\nu-1} \Delta^{\nu-1} \left[\frac{N}{x} \log \frac{N}{N-x} \right]_{x=0} . \quad \dots \quad (6.4) \end{aligned}$$

where, for $x = 0$, $\frac{N}{x} \log \frac{N}{N-x}$ is to be interpreted as 1.

If $\bar{y}^*$ be the average Y-characteristics of the ν distinct units then

$$V(\bar{y}^*) = \frac{N-\nu}{N-1} \frac{\sigma^2}{\nu} . \quad \dots \quad (6.5)$$

That (6.5) is smaller than (6.4) may be proved in precisely the same way as we proved a similar result in §5.

7. SAMPLING WITH UNEQUAL PROBABILITIES

We consider now the more general situation where sampling is done with different probabilities attached to the different population units. Let P_j be the probability associated with the j-th population unit ($\Sigma P_j = 1$). Suppose a sample of size n is taken with replacement and with the P_j's as the probabilities. Let us suppose that for the i-th sample unit we record its Y-characteristic y_i, its probability of selection p_i, and its unit-index $u_i (i = 1, 2, \ldots, n)$. Thus, the sample is

$$\mathbf{S} = (\mathbf{x}_1, \mathbf{x}_2 \ldots, \mathbf{x}_n)$$

where

$$\mathbf{x}_i = (y_i, p_i, u_i) \quad (i = 1, 2, \ldots, n).$$

Clearly the $\mathbf{x}_i$'s are independently and identically distributed random vectors.

As in §5 let us define ν as the number of distinct x_i's and as before let $\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots \mathbf{x}_{(\nu)}$ be an arrangement of the ν distinct $\mathbf{x}_i$'s in ascending order of their unit-indices. Let

$$\mathbf{T} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(\nu)}).$$

It is easily seen that in this case also $\mathbf{T}$ is a sufficient statistic. Given $\mathbf{T}$, there are only

$$n! \ \Sigma' \ (\alpha_1! \ \ \alpha_2! \ \ \ldots. \ \ \alpha_\nu!)^{-1}$$

(where the summation is taken over all positive integer α_i's such that $\alpha_1 + \ldots + \alpha_\nu = n$) values that $\mathbf{S}$ may take and the probability for each of these can be computed from the information $\mathbf{T}$ alone. Thus, any admissible estimator of $\bar{Y}$ must necessarily be a function of the statistic $\mathbf{T}$ alone. The usual estimator $\bar{y}$ of $\bar{Y}$ is based on all the n observations and is

$$\bar{y} = \bar{y}(\mathbf{S}) = \frac{1}{n} \sum_{1}^{n} \left(\frac{y_i}{Np_i} \right). \qquad \ldots \quad (7.1)$$

From the sufficiency of $\mathbf{T}$ it follows that

$$\bar{y}^* = E(\bar{y} \,|\, \mathbf{T}) = E\left(\frac{y_1}{Np_1} \middle| \mathbf{T} \right) \qquad \ldots \quad (7.2)$$

is a better unbiased estimator of $\bar{Y}$.

It is not difficult to show that

$$P(\mathbf{x}_1 = \mathbf{x}_{(i)} \,|\, \mathbf{T}) = \frac{p_{(i)} \sum'' \dfrac{(n-1)!}{\alpha_1! \ \alpha_2! \ldots \alpha_\nu!} \, p_{(1)}^{\alpha_1} \, p_{(2)}^{\alpha_2} \cdots p_{(\nu)}^{\alpha_\nu}}{\sum' \dfrac{n!}{\alpha_1! \alpha_2! \ldots \alpha_\nu!} \, p_{(1)}^{\alpha_1} \, p_{(2)}^{\alpha_2} \cdots p_{(\nu)}^{\alpha_\nu}}$$

$$= c_i (\text{say}) \quad (i = 1, 2, \ldots, \nu)$$

where Σ' means summation over all integral α's such that

$$\alpha_1 + \alpha_2 + \ldots + \alpha_\nu = n \ \text{ and } \ \alpha_k > 0 \ \text{ for } \ k = 1, 2, \ldots, \nu$$

and Σ'' means summation over all integral α's such that

$$\alpha_1 + \alpha_2 + \ldots + \alpha_\nu = n-1, \ \alpha_i \geqslant 0 \text{ and } \alpha_k > 0 \text{ for } k \neq i.$$

Thus

$$\bar{y}^* = E\left\{ \frac{y_1}{Np_1} \middle| \mathbf{T} \right\} = \sum_{i} c_i \frac{y_{(i)}}{Np_{(i)}}. \qquad \ldots \quad (7.3)$$

Unfortunately, it is rather troublesome to compute the c_i's. In the particular case where $n = 3$ and $\nu = 2$ we have

$$c_1 = (2p_{(1)} + p_{(2)})/3(p_{(1)} + p_{(2)})$$

and

$$c_2 = (p_{(1)} + 2p_{(2)})/3(p_{(1)} + p_{(2)}).$$

The estimator $\bar{y}^*$, though demonstrated to be superior to the usual estimator $\bar{y}$, cannot be of much use for large scale sample surveys. It is even more troublesome to estimate the variance of $\bar{y}^*$. An unbiased estimator for the variance of $\bar{y}$ is

$$\frac{1}{n(n-1)} \sum_{1}^{n} \left(\frac{y_i}{Np_i} - \bar{y} \right)^2. \qquad \ldots \quad (7.4)$$

The above will over-estimate the variance of $\bar{y}^*$ and so it will be on the safe side to take (7.4) as an estimator of the variance of $\bar{y}^*$.

8. THE MEAN AND VARIANCE OF ν

In § 2 we have given the distribution of ν for the particular case where sampling is done with equal probabilities. In the unequal probability set-up the distribution of ν becomes very messy. Here we give expressions for the mean and variance of ν.

Let z_j be the characteristic function of the event that the sample of n units includes the j-th population unit.

Clearly

$$P(z_j = 1) = 1 - Q_j^n \quad (j = 1, 2, \ldots, N)$$

where

$$Q_j = 1 - P_j.$$

Since

$$\nu = z_1 + z_2 + \ldots, + z_N \qquad \ldots \quad (8.1)$$

we have

$$E(\nu) = \sum_{1}^{N} (1 - Q_j^n). \qquad \ldots \quad (8.2)$$

Also

$$V(\nu) = \Sigma\, V(z_j) + \sum_{i \neq j} \text{cov}\ (z_i, z_j). \qquad \ldots \quad (8.3)$$

Now,

$$V(z_j) = Q_j^n (1 - Q_j^n)$$

and

$$\begin{aligned} \text{cov}(z_i, z_j) &= P(z_i = z_j = 1) - P(z_i = 1) P(z_j = 1) \\ &= (1 - Q_i^n - Q_j^n + Q_{ij}^n) - (1 - Q_i^n)(1 - Q_j^n) \\ &= -(Q_i^n\ Q_j^n - Q_{ij}^n) \end{aligned}$$

where

$$Q_{ij} = 1 - P_i - P_j$$

$$\therefore \quad V(\nu) = \Sigma\ Q^n(1 - Q^n) - \sum_{i \neq j} (Q^n Q^n - Q_{ij}^n) = \Sigma\ Q_i^n - (\Sigma\, Q_i^n)^2 + \sum_{i \neq j} Q_{ij}^n. \qquad \ldots \quad (8.4)$$

9. UNEQUAL PROBABILITIES AND WITHOUT REPLACEMENT

Now let us consider the case of sampling without replacement and with different probabilities. As in §7 let P_j be the probability attached to the j-th population unit. Let y_i and p_i be the Y-characteristic and the probability corresponding to the i-th sample unit $(i = 1, 2, \ldots, n)$. Writing $\mathbf{x}_i = (y_i, p_i)$ we can record the sample observation[1] as

$$\mathbf{S} = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n). \qquad \ldots \quad (9.1)$$

---

[1] Here we need not record the unit indices of the sample units as we know that they must be all different. Unless we have some additional information about the population units it appears that it is impossible to utilise any information about the sample unit-indices to improve on any estimator of $\bar{Y}$.

Now, let us order the $\mathbf{x}_i$'s by some method, say, in ascending order of the y_i's and for $\mathbf{x}_i$'s with equal y_i's, in ascending order of their p_i's. Let $\mathbf{x}_{(i)}$ be the i-th order statistic and let

$$\mathbf{T} = (\mathbf{x}_{(1)}, \mathbf{x}_{(2)}, \ldots, \mathbf{x}_{(n)}) \qquad \ldots \quad (9.2)$$

be the set of order statistics.

For given $\mathbf{T}$, there are $n!$ or less (in the case where some of the $\mathbf{x}_{(i)}$'s are the same) values that $\mathbf{S}$ may take and the conditional probability for each of these may be computed from the information $\mathbf{T}$ alone. Thus, in this case also $\mathbf{T}$ is a sufficient statistic. Hence no estimator that is not a function of $\mathbf{T}$ alone can be admissible. Any estimator that makes use of the order in which the sample was drawn can be uniformly improved upon by its conditional expectation given $\mathbf{T}$.

For example, consider the particular case of $n = 2$. We may estimate $\bar{Y}$ from the first component of $\mathbf{S}$, i.e., from $\mathbf{x}_1$. This estimator is, of course, $y_1 | Np_1$.

Now
$$P(\mathbf{x}_1 = \mathbf{x}_{(1)} | \mathbf{T}) = \frac{p_{(1)}\, p_{(2)}/(1-p_{(1)})}{p_{(1)}\, p_{(2)}/(1-p_{(1)})+p_{(2)}\, p_{(1)}/(1-p_{(2)})} = \frac{1-p_{(2)}}{2-p_{(1)}-p_{(2)}} \quad \ldots \quad (9.3)$$

and similarly

$$P(\mathbf{x}_1 = \mathbf{x}_{(2)} | \mathbf{T}) = \frac{1-p_{(1)}}{2-p_{(1)}-p_{(2)}} \qquad \ldots \quad (9.4)$$

$$\therefore \; E\left(\frac{y_1}{Np_1}\middle| \mathbf{T}\right) = \left[(1-p_{(2)})\ \frac{y_{(1)}}{Np_{(1)}} + (1-p_{(1)})\frac{y_{(2)}}{\mathrm{t}Np_{(2)}}\right] \times$$
$$\times (2-p_{(1)}-p_{(2)})^{-1}. \qquad \ldots \quad (9.5)$$

The estimator (9.5) is uniformly better than $y_1 | Np_1$.

Since $\mathbf{T}$ is not a complete sufficient statistic, we cannot prove that (9.5) is the uniformly best estimator. For further discussion about the problem dealt with in this section one may refer to (Murthy, 1957).

Acknowledgement

The author wishes to thank Mr. S. Raja Rao who suggested the problem in 1952 and then collaborated with the author in obtaining the results contained in § 2 to § 6.

References

Feller, W. (1957) : *An introduction to probability theory and its applications*, John Wiley & Sons New York.

Fraser, D. A. S. (1957) : *Non parametric methods in statistics*, John Wiley & Sons New York.

Murthy, M. N. (1957) : Ordered and unordered estimators in sampling without replacement. *Sankhyā*, **18**, 379.

Paper received : January, 1958.

DIFFUSION BY DISCRETE MOVEMENTS

By H. C. GUPTA
University of Delhi

SUMMARY. The problem of diffusion or migration in a straight line by discontinuous movements was discussed briefly by Taylor (1922) and a more complete investigation was presented by Goldstein (1951). This problem has been studied in somewhat greater generality in Part I of this paper; in Part II the effect of barriers (reflecting, or absorbing, or partly reflecting and partly absorbing), has been investigated and amongst the formulae obtained here some are generalizations of certain results given by Chandrasekhar (1943).

1. Introduction

Suppose that a number of non-interacting particles start from the origin at the time $t = 0$ with the velocities u_1 (> 0) and u_2(< 0) in the proportions c_1 and c_2 ($= 1-c_1$) respectively and move along a straight line for an interval of time τ. After time τ and at the end of each interval τ thereafter, a particle has a probability p of maintaining its motion unchanged and a probability q ($= 1-p$) of changing over to the other possible velocity. Thus the matrix of the transition probabilities is

$$\begin{array}{ccc} \text{velocity} & \multicolumn{2}{c}{\text{from}} \\ & u_1 & u_2 \\ \text{to} & & \\ u_1 & \left[\begin{array}{c} p \\ q \end{array}\right. & \left.\begin{array}{c} q \\ p \end{array}\right] \\ u_2 & & \end{array} \qquad \dots \quad (1.1)$$

To correspond to physical reality a large number of non-interacting particles should be supposed to be moving along a straight line and the various probability statements are to be construed in terms of the density of the particles. This remark is of special relevance to § 5, where the possibility of emigration and immigration is also included. The diffusion process considered here is Markovian in the sense that if the realization of the process at the time t is known, then the knowledge of the realizations before time t is of no relevance to the process after time t; and the history of any given particle may be characterized as a Markov process with two states.

2. The characteristic functions (c.f.) of the distributions

Let $\alpha(n, \nu)$ denote the probability that a particle has at the time $n\tau-0$ the velocity u_1, that is, at the end of the n-th interval before a transition, if any, has taken place, the particle arrives from the left after having moved to the right during $\frac{1}{2}(n+\nu)$ time-intervals and to the left during $\frac{1}{2}(n-\nu)$ intervals, so that during the last, i.e., n-th interval, its velocity has been u_1 and at the end of this interval its distance from the origin is

$$\tfrac{1}{2}\tau\{u_1(n+\nu)+u_2(n-\nu)\}. \qquad \dots \quad (2.1)$$

Similarly define $\beta(n, \nu)$ as the probability that the particle has at time $n\tau-0$ the velocity u_2, after having moved during $\frac{1}{2}(n+\nu)$ intervals to the right and during $\frac{1}{2}(n-\nu)$ intervals to the left.

Evidently n is a positive integer and ν is an integer such that

$$\alpha(n, \nu) = 0 = \beta(n, \nu), \text{ if } n-\nu \text{ is odd, or if } |\nu| > n. \qquad \ldots \quad (2.2)$$

Elementary reasoning shows that for $n \geqslant 1$,

$$\alpha(n, -n) = 0 = \beta(n, n), \alpha(n, n) = c_1 p^{n-1} \text{ and } \beta(n, -n) = c_2 p^{n-1};$$

also,
$$\alpha(2, 0) = c_2 q \text{ and } \beta(2, 0) = c_1 q. \qquad \ldots \quad (2.3)$$

In general, the following recurrence relations hold:

$$\left.\begin{aligned} \alpha(n+1, \nu) &= p\alpha(n, \nu-1)+q\beta(n, \nu-1), \\ \beta(n+1, \nu) &= p\beta(n,\nu+1)+q\alpha(n,\nu+1). \end{aligned}\right\} \qquad \ldots \quad (2.4)$$

Now define the characteristic functions (c.f.)

$$L_n(\theta) = \sum_{\nu=-n}^{n} \alpha(n, \nu)e^{i\nu\theta} \; ; \; M_n(\theta) = \sum_{\nu=-n}^{n} \beta(n, \nu)e^{i\nu\theta},$$

where θ is a real number. In fact, on account of the relations (2.2) and (2.3),

$$L_n = \sum_{\nu=-n+2}^{n} \text{ and } M_n = \sum_{\nu=-n}^{n-2}.$$

From (2.3),

$$L_1 = c_1 e^{i\theta}, \quad M_1 = c_2 e^{-i\theta}, \quad L_2 = c_2 q + c_1 p e^{2i\theta}, \quad M_2 = c_2 p e^{-2i\theta} + c_1 q. \qquad \ldots \quad (2.5)$$

Multiply (2.4) by $e^{i\nu\theta}$ and sum from $\nu = -(n+1)$ to $n+1$; then by (2.2)

$$L_{n+1}(\theta) = pe^{i\theta} \sum_{\nu-1=-n}^{n} \alpha(n, \nu-1)e^{i(\nu-1)\theta} + qe^{i\theta} \sum_{\nu-1=-n}^{n} \beta(n, \nu-1)e^{i(\nu-1)\theta},$$

$$M_{n+1}(\theta) = pe^{-i\theta} \sum_{\nu+1=-n}^{n} \beta(n, \nu+1)e^{i(\nu+1)\theta} + qe^{-i\theta} \sum_{\nu+1=-n}^{n} \alpha(n, \nu+1)e^{i(\nu+1)\theta}.$$

Denoting by E the operator which increases the suffix of L_n or M_n by 1, these relations may be written

$$(E-pe^{i\theta})\, L_n(\theta) = qe^{i\theta} M_n(\theta) \quad \text{and} \quad (E-pe^{-i\theta})\, M_n(\theta) = qe^{-i\theta}\, L_n(\theta), \quad \ldots \quad (2.6)$$

showing that both L_n and M_n satisfy the difference equation

$$(E-pe^{i\theta})(E-pe^{-i\theta})\chi_n(\theta) = q^2\chi_n(\theta), \text{ i.e., } \chi_{n+2} - 2p\cos\theta\, \chi_{n+1} + (p^2-q^2)\chi_n = 0,$$

whence
$$M_n(\theta), L_n(\theta) = \tfrac{1}{2}k_1(a_1^n + a_2^n) + \tfrac{1}{2}k_2(a_1^n - a_2^n), \qquad \ldots \quad (2.7)$$

where a_1 and a_2 are the roots of the quadratic $x^2-2p\cos\theta . x+p^2-q^2 = 0$, i.e., with

$$\left.\begin{aligned} r = \frac{p}{q},\ p = \frac{r}{r+1},\ q = \frac{1}{r+1} \text{ and } c \equiv p-q = \frac{r-1}{r+1}, \\ a_1 = p\cos\theta+q\surd(1-r^2\sin^2\theta) \quad\text{and}\quad a_2 = p\cos\theta-q\surd(1-r^2\sin^2\theta). \end{aligned}\right\} \quad \ldots \quad (2.8)$$

It will be supposed that the radical takes the positive sign throughout. The constants k_1 and k_2 have to be consistent with the values (2.5). Thus for L_n they are given by

$$c_1e^{i\theta} = k_1p\cos\theta+k_2q(1-r^2\sin^2\theta)^{\frac{1}{2}},$$

and
$$c_2q+c_1pe^{2i\theta} = k_1(p^2\cos 2\theta+q^2)+2pq\cos\theta . k_2\,(1-r^2\sin^2\theta)^{\frac{1}{2}},$$

whence, unless $p = q$,

$$k_1 = \frac{c_1p-c_2q}{p-q} = \frac{c_1-q}{c} \text{ and } k_2 = \frac{ic_1\sin\theta\, p^2-c_1q^2e^{i\theta}+c_2pq\cos\theta}{(p-q)q(1-r^2\sin^2\theta)^{\frac{1}{2}}}. \quad \ldots \quad (2.9)$$

The expression for L_n can now be written from (2.7) and (2.9). Similarly it is found that

$$M_n(\theta) = \frac{c_2p-c_1q}{2(p-q)}\,(a_1^n+a_2^n)+\frac{-ic_2p^2\sin\theta-c_2q^2e^{-i\theta}+c_1pq\cos\theta}{2q(p-q)(1-r^2\sin^2\theta)^{\frac{1}{2}}}(a_1^n-a_2^n), \quad \ldots \quad (2.10)$$

so that
$$L_n+M_n = \tfrac{1}{2}(a_1^n+a_2^n)+\frac{i\sin\theta(c_1-c_2)+q\cos\theta}{2q(1-r^2\sin^2\theta)^{\frac{1}{2}}}\,(a_1^n-a_2^n),$$

which in the special case $c_1 = c_2 = \frac{1}{2}$ reduces to the expression G(16) (Goldstein, 1951) taken along with G(19) (1951).[1]

In the case $p = q$, an application of the theory of limits gives

$$L_1(\theta) = c_1e^{i\theta},\ L_n(\theta) = \tfrac{1}{2}\cos^{n-2}\theta(c_2+c_1e^{2i\theta}),\ n > 1;$$

$$M_1(\theta) = c_2e^{-i\theta},\ M_n(\theta) = \tfrac{1}{2}\cos^{n-2}\theta(c_1+c_2e^{-2i\theta}),\ n > 1.$$

3. The probability distributions

On using the analogue of the Fourier inversion formula for the discrete case, the relation (2.7) gives

$$\alpha(n, \nu) = \frac{1}{4\pi}\int_0^{2\pi} e^{-i\nu\theta}[k_1(a_1^n+a_2^n)+k_2(a_1^n-a_2^n)]\,d\theta.$$

---

[1] The equations occurring in Goldstein's paper are prefixed by G.

Since $a_r(\theta) = a_r(2\pi-\theta)$, $r = 1, 2$ and k_1 is independent of θ, it is easy to see that

$$\alpha(n, \nu) = \frac{1}{2\pi}\int_0^\pi [\cos \nu\theta\{k_1(a_1^n+a_2^n)+(c_2p-c_1q)\cos\theta(a_1^n-a_2^n)/c\sqrt{(1-r^2\sin^2\theta)}\}+$$
$$+\frac{c_1\sin\theta\sin\nu\theta}{q\sqrt{(1-r^2\sin^2\theta)}}(a_1^n-a_2^n)]\,d\theta. \quad \ldots \quad (3.1)$$

It may be verified that $\alpha(n, \nu) = 0$ when $n-\nu$ is odd, for from (2.8)

$$a_1(\pi-\theta) = -a_2(\theta) \text{ and } a_2(\pi-\theta) = -a_1(\theta),$$

so that
$$a_1^n(\pi-\theta)+a_2^n(\pi-\theta) = (-)^n(a_1^n+a_2^n),$$

while
$$a_1^n(\pi-\theta)-a_2^n(\pi-\theta) = (-)^{n-1}(a_1^n-a_2^n);$$

and the integrand in (3.1), on replacing θ by $\pi-\theta$, becomes $(-1)^{n+\nu}$, i.e., -1, times its previous value, showing that $\alpha(n,\nu) = 0$. This shows further that when $n+\nu$ is even,

$$\alpha(n\ \nu) = \frac{1}{\pi}\int_0^{\frac{1}{2}\pi}\left\{ k_1(a_1^n+a_2^n)\cos\nu\theta+\frac{a_1^n-a_2^n}{\sqrt{1-r^2\sin^2\theta}}\left(\frac{p-c_1}{c}\cos\theta\cos\nu\theta+\left(\frac{c_1}{q}\right)\sin\theta\sin\nu\theta\right)\right\}d\theta. \quad \ldots \quad (3.2)$$

Making the substitution $\theta = \pi-\phi$ in the part of the integrand multiplied by a_2^n, so that the new limits for this part become $\frac{1}{2}\pi$ and π, and combining the two parts, we have

$$\alpha(n, \nu) = \frac{1}{\pi}\int_0^\pi \left\{ k_1\cos\nu\theta+\frac{l_1\cos(\nu-1)\theta+m_1\cos(\nu+1)\theta}{2\sqrt{(1-r^2\sin^2\theta)}}\right\} a_1^n\,d\theta, \quad \ldots \quad (3.3)$$

where by (2.9), $k_1 = \dfrac{c_1-q}{c}$, $l_1 = \dfrac{p-c_1}{c}+\dfrac{c_1}{q}$ and $m_1 = \dfrac{p-c_1}{c}-\dfrac{c_1}{q}$, ... (3.4)

Similarly
$$\beta(n, \nu) = \frac{1}{\pi}\int_0^\pi \left\{ (1-k_1)\cos\nu\theta+\frac{l_2\cos(\nu+1)\theta+m_2\cos(\nu-1)\theta}{2\sqrt{(1-r^2\sin^2\theta)}}\right\} a_1^n\,d\theta, \quad \ldots \quad (3.5)$$

where
$$l_2 = k_1+c_2/q \text{ and } m_2 = k_1-c_2/q. \quad \ldots \quad (3.6)$$

Since
$$\tfrac{1}{2}(l_1-m_1-l_2+m_2) = (c_1-c_2)/q \text{ and } \tfrac{1}{2}(l_1+m_1+l_2+m_2) = 1,$$

we see that

$$\gamma(n, \nu) \equiv \alpha(n, \nu)+\beta(n, \nu)$$

$$= \frac{1}{\pi}\int_0^\pi a_1^n\,[\cos\nu\theta\left(1+\frac{\cos\theta}{\sqrt{(1-r^2\sin^2\theta)}}\right)+\frac{c_1-c_2}{q}\,\frac{\sin\nu\theta\sin\theta}{\sqrt{(1-r^2\sin^2\theta)}}\Big\}\,d\theta,$$

which agrees with the integral G(26) (1951) for $c_1 = c_2$.

4. THE MOMENTS OF THE DISTRIBUTIONS

Remembering that the coordinate ν in $\alpha(n, \nu)$ or in $\beta(n, \nu)$ corresponds to a distance from the origin equal to

$$\tfrac{1}{2}\tau\{(u_1+u_2)n+(u_1-u_2)\nu\} = (nu_0+\nu u)\tau = (m+\nu)d, \qquad \dots \quad (4.1)$$

where
$$u_0 = \tfrac{1}{2}(u_1+u_2), \quad u = \tfrac{1}{2}(u_1-u_2), \quad m = nu_0/u \quad \text{and} \quad d = u\tau, \qquad \dots \quad (4.2)$$

the r-th moment about the origin for the α-distribution is

$$d^r \sum_{\nu=-n}^{n} (\nu+m)^r \, \alpha(n, \nu)/A, \qquad \dots \quad (4.3)$$

where A is the probability mass

$$\sum_{\nu=-n}^{n} \alpha(n, \nu) = L_n(0) \; = \; \tfrac{1}{2}\{1+(p-q)^n\} \, \frac{c_1p-c_2q}{p-q} + \tfrac{1}{2}\{1-(p-q)^n\} \, \frac{c_2p-c_1q}{p-q}$$

$$= \; \tfrac{1}{2}+\tfrac{1}{2}(c_1-c_2)(p-q)^{n-1}, \qquad \dots \quad (4.4)$$

so that $\alpha(n, \nu)/A$ is a proper probability distribution.

Since $|p-q| < 1$, it is seen that $\lim\limits_{n \to \infty} A = \tfrac{1}{2}$, showing that the probability of a particle moving to the right is ultimately $\tfrac{1}{2}$. Thus there is a tendency to an equilibrium state with regard to the velocity. This is in agreement with the important property (Feller, 1953) of an irreducible finite aperiodic chain that the absolute probabilities of the states have a limiting stationary distribution and if the chain is governed by a doubly stochastic matrix of transition probabilities like (1.1), this stationary distribution has each element equal to $1/n$ (Bartlett, 1955), n being the number of possible states.

To compute (4.3) we observe that $\Sigma \, \nu^r \alpha(n, \nu) = (-i)^r \, r!$ times the coefficient of θ^r in the power-series expansion of $L_n(\theta)$.

Thus
$$\sum_{\nu=-n}^{n} \nu\alpha(n,\nu) = \tfrac{1}{2}\{1-(p-q)^n\}c_1/q, \qquad \dots \quad (4.5)$$

and
$$\sum_{\nu=-n}^{n} \nu^2\alpha(n, \nu) = \frac{c_1p-c_2q}{p-q} \cdot \frac{nr}{2}\Big\{1-(p-q)^n\Big\} + \frac{c_2p-c_1q}{p-q} \, \frac{nr}{2}\Big\{1+(p-q)^n\Big\} +$$

$$+(1-r^2)(c_1q-c_2p)\{1-(p-q)^n\}/2(q-p)$$

$$= \tfrac{1}{2}nr\{1+(c_2-c_1)(p-q)^{n-1}\} + \frac{1}{2q^2} \, (c_1q-c_2p)\{1-(p-q)^n\}. \qquad \dots \quad (4.6)$$

Hence using the results (4.3) to (4.6), the distances from the origin of the particles moving to the right (with the velocity u_1) at the end of the n-th interval have

$$\text{mean} = d(1-c^n)c_1/q\{1+(c_1-c_2)c^{n-1}\}+md,$$

and mean-square

$$= m^2d^2 + \frac{d^2}{1+(c_1-c_2)c^{n-1}} \left[nr\{1+(c_2-c_1)c^{n-1}\} + \frac{c_1q-c_2p}{q^2} \, (1-c^n) + \frac{2mc_1(1-c^n)}{q} \right]$$

$$= d^2(m^2-nr) + \frac{d^2[2np+(1-c^n)(c_1-c_2r+2mc_1)]}{q\{1+(c_1-c_2)c^{n-1}\}}.$$

A comparison of the integrals (3.3) and (3.5) would reveal, what is clear from a little consideration, that $\beta(n, \nu)$ may be obtained from $\alpha(n, \nu)$ by interchanging the suffixes 1, 2 and replacing ν by $-\nu$. Consequently u, m, d replaced respectively by $-u$, $-m$, $-d$ in the r-th moment of the α-distribution give the r-th moment of the β-distribution. Using these moments, the moments of the distance unconditional on the velocity may be derived. In particular, we have

$$\left.\begin{aligned} \text{mean} &= md+(c_1-c_2)\{1-(p-q)^n\}d/2q, \\ \text{and} \qquad \text{mean-square} &= d^2[m^2+nr+(1-c^n)\{\tfrac{1}{2}(1-r^2)+m(c_1-c_2)/q\}]. \end{aligned}\right\} \quad \ldots \quad (4.7)$$

The variances of the distributions could now be calculated, but the main interest lies in the limits of the various expressions when the discrete movements become continuous, i.e., when $\tau \to 0$ and $n \to \infty$ such that $n\tau \to t$ and $q/\tau \to Q$; then $m \to \infty$, $d \to 0$ but $md \to u_0 t$. Making these substitutions and noting that

$$c^n = (p-q)^n = (1-2Q\tau)^n \sim e^{-2Qt},$$

we arrive at the expressions

$$\left.\begin{aligned} \bar{x} &= u_0 t+\tfrac{1}{2}(c_1-c_2)(1-e^{-2Qt})u/Q, \\ \text{and} \qquad \overline{x^2} &= u_0^2 t^2+u^2 t/Q+(c_1-c_2)uu_0 t/Q-u^2/2Q^2+e^{-2Qt}\{u^2/2Q^2+(c_2-c_1)uu_0 t/Q\}. \end{aligned}\right\} \ldots \quad (4.8)$$

In the special case $c_1 = \frac{1}{2} = c_2$ and $u_1 = -u_2 = u$, so that $u_0 = 0$, $\bar{x} = 0$, as it should and

$$\overline{x^2} = u^2 t/Q-(1-e^{-2Qt})u^2/2Q^2, \quad \ldots \quad (4.9)$$

which is in agreement with the value G(2) obtained by Taylor (1922).

5. MODIFICATIONS FOR IMMIGRATION AND EMIGRATION

Next consider the generalization when $p+q \neq 1$. If $p+q < 1$, then at the end of a partial path there is the probability $1-p-q$ of the particle disappearing, and if $p+q > 1$, then there is the probability $p+q-1$ that another particle having the same position and velocity as the original joins. Consequently we dispense with the condition $p+q = 1$ and as before think in terms of a large number of particles participating in the motion. The equations upto and including (2.7) remain unaltered with a_1 and a_2 still given by (2.8), but k_1 and k_2 have values as given in (2.9) multiplied by $1/(p+q)$ and so have $L_n(\theta)$ and $M_n(\theta)$. The new expressions for the c.f.'s L_n and M_n become homogeneous in p and q, so also those of the various moments. The expressions for the mean distance and the mean-square distance from the origin also become homogeneous, in fact of zero degree in p, q; the latter for particles moving to the right at the end of the n-th partial path is

$$(m^2-nr)d^2+\frac{d^2}{q}\,\frac{2np(p+q)^{n-1}+\{(p+q)^n-(p-q)^n\}(c_1-c_2 r+2mc_1)}{(p+q)^{n-1}+(c_1-c_2)(p-q)^{n-1}}, \quad \ldots \quad (5.1)$$

while the mean-square distance unconditional on the velocity is

$$d^2[m^2+nr+(1+r)\{1-(p-q)^n/(p+q)^n\}\{\tfrac{1}{2}(1-r)+m(c_1-c_2)\}]. \quad \ldots \quad (5.2)$$

Proceeding now to the limits as before but with $p = 1-P\tau$, the limiting values of the moments are found to be $e^{(Q-P)t}$ times those obtained in § 4, but those for the mean distances and mean-square distances are the same as in § 4. It is thus seen that emigration here has no effect on the mean and the mean-square distances. It is what is to be expected, since emigration takes place at a constant rate for all x and for the two velocities.

6. DIFFERENTIAL EQUATIONS GOVERNING CONTINUOUS DIFFUSION

We now study how by limiting processes the various differential equations, for diffusion, etc. may be obtained from the difference equation for $\alpha(n, \nu)$ obtained from (2.4) by elimination of $\beta(n, \nu)$, namely from the equation

$$\alpha(n+1, \nu)-p\{\alpha(n, \nu-1) + \alpha(n, \nu+1)\}+(p^2-q^2)\alpha(n-1, \nu) = 0. \qquad \dots \quad (6.1)$$

In the coordinates (n, ν), n is to be identified with time and ν with position by means of (2.1), i.e., n corresponds to $n\tau = T$ and ν to

$$\tfrac{1}{2}(n+\nu)u_1\tau+\tfrac{1}{2}(n-\nu)u_2\tau = u_0T+\nu u\tau = y, \text{ say.} \qquad \dots \quad (6.2)$$

Since in the discrete case the particles, for a given n, can move to such positions only that ν differs by 2, we must in the continuous case take $2u\tau = \delta$ as the differential of the distance from the origin; thus (6.1) may equivalently be written

$$\alpha(T+\tau, y)-p\{\alpha(T, y-\tfrac{1}{2}\delta)+\alpha(T, y+\tfrac{1}{2}\delta)\}+(p^2-q^2)\,\alpha(T-\tau, y) = 0.$$

Expand in Taylor's series, neglecting terms of a higher order than τ^2 or δ^2; then we have

$$(1-2p+p^2-q^2)\alpha+(1-p^2+q^2)\tau\frac{\partial\alpha}{\partial T}+\tfrac{1}{2}\tau^2(1-p^2-q^2)\,\frac{\partial^2\alpha}{\partial T^2}-\tfrac{1}{4}p\delta^2\frac{\partial^2\alpha}{\partial y^2} = 0. \qquad \dots \quad (6.3)$$

It is to be noted that y involves T also; we, therefore, transform the variables from T, y to t, x by means of the relations

$$t = T, \quad x = y-u_0T,$$

and (6.3) transforms to

$$(1-p-q)(1-p+q)\alpha+\tau(1-p^2+q^2)\left(\frac{\partial\alpha}{\partial t}-u_0\frac{\partial\alpha}{\partial x}\right)+$$
$$+\tfrac{1}{2}\tau^2(1+p^2-q^2)\left\{\frac{\partial^2\alpha}{\partial t^2}-2u_0\frac{\partial^2\alpha}{\partial x\partial t}+u_0^2\frac{\partial^2\alpha}{\partial x^2}\right\}-\tfrac{1}{4}p\delta^2\frac{\partial^2\alpha}{\partial x^2} = 0. \qquad \dots \quad (6.4)$$

The following limiting operations are of interest :

(i) $p = q = \tfrac{1}{2}$, $2u\tau = \delta = O(\sqrt{\tau})$, so that $u = O(1/\sqrt{\tau}) = \sqrt{(2k/\tau)}$, say; then (6.4) reduces to

$$\frac{\partial\alpha}{\partial t}-u_0\frac{\partial\alpha}{\partial x}-k\frac{\partial^2\alpha}{\partial x^2} = 0, \qquad \dots \quad (6.5)$$

which is the equation of heat conduction if $u_0 = 0$, k being the thermometric conductivity; it is also the equation for the probability density function $\alpha(x)$ for one-dimensional diffusion from a point source with a constant drift in the random displacement X at rate $-u_o$ per unit time and a growth in variance at rate $2k$.

(ii) $p = q = \frac{1}{2}$, $\delta = O(\tau^{\frac{1}{2}+\varepsilon})$, $\epsilon > 0$; then (6.4) reduces to

$$\frac{\partial \alpha}{\partial t} = u_0 \frac{\partial \alpha}{\partial x}, \qquad \dots \quad (6.6)$$

which is the equation of the motion of a sound wave. It will be noticed that in the above two cases the possible velocities u_1 and u_2 are infinite, but their algebraic sum $2u_0$ is finite. In the following cases where $\delta = O(\tau)$, the velocities, with which diffusion is possible, are both finite.

(iii) $p = 1-P\tau$, $q = Q\tau$, $[P, Q$ finite$]$, $\delta = 2u\tau = O(\tau)$ so that u is finite; then on dividing by τ^2 and letting $\tau \to 0$, (6.4) becomes

$$\left(\frac{\partial}{\partial t} - u_1 \frac{\partial}{\partial x} + P\right)\left(\frac{\partial}{\partial t} - u_2 \frac{\partial}{\partial x} + P\right) \alpha(x, t) = Q^2 \alpha(x, t). \qquad \dots \quad (6.7)$$

This is the equation of diffusion by continuous movements, with two possible velocities $u_1 (> 0)$ and $u_2 (< 0)$, a constant rate of transition Q and a constant rate of emigration $P-Q$. When $u_1 = -u_2 = v$, then (6.7) reduces to the telegraph equation with leakage ; its solution[1] for the case $P = Q$ and $c_1 = c_2 = \frac{1}{2}$ has been obtained by Goldstein (1951). By the method of the Fourier and Laplace transformations, the writer has obtained the solution[1] of a still more general equation, viz.

$$\left\{\left(\frac{\partial}{\partial t} - u_1 \frac{\partial}{\partial x} + P_1\right)\left(\frac{\partial}{\partial t} - u_2 \frac{\partial}{\partial x} + P_2\right) - Q^2\right\} \alpha(x, t) = 0.$$

Although the limiting processes (i) and (iii) are essentially different (the possible velocities being infinite in the one and finite in the other), still it is of some interest to compare the solutions for the probability density in the two cases. It will be found that for t finite and x/v small, so that x is finite in case (i) but small in case (iii), and a large rate of transition Q (in fact, $Q\tau = 1$, τ being the small time interval in the discrete case) the two solutions do agree. The demonstration is as follows :

The probability density for the random walk problem (Chandrasekhar, 1943) is

$$W(x, t) = \frac{1}{2\sqrt{(\pi D t)}} \exp\left(-\frac{x^2}{4Dt}\right), \; N \gg m, \qquad \dots \quad (6.8)$$

where n' = number of displacements per unit time, $D = \frac{1}{2}n'l^2$, l = length of a step $= v\tau$, v = velocity of diffusion, $ml = x$ = the net displacement from the starting point, $n't$ = N = the number of steps taken in time t. Thus $vt = Nl$, $n' = 1/\tau$ and (6.8) may be written

$$\chi(x, t) = \frac{1}{v\sqrt{(2\pi t\tau)}} \exp\left(-\frac{x^2}{2v^2 t\tau}\right), \; x \ll vt. \qquad \dots \quad (6.9)$$

---

[1] Under publication.

Now Goldstein's solution (1951), leaving out the discrete probability mass at the outposts of the distribution, is

$$\chi(x, t) = \frac{Q}{2v} e^{-Qt} \left[I_0(z) + \frac{Qt}{z} I_1(z) \right], \; z \equiv Q\sqrt{\left(t^2 - \frac{x^2}{v^2} \right)}. \qquad \ldots \; (6.10)$$

It will now be shown that (6.9) and (6.10) agree to a first approximation, if, τ being a small time interval,

$$Q = 1/\tau, \quad t = O(1) \quad \text{and} \quad x\sqrt{Q}/v = O(1). \qquad \ldots \; (6.11)$$

For under these assumptions

$$z \sim Qt\left(1 - \frac{x^2}{2v^2t^2} \right) = O(Q), \text{ so that } I_0(z) \sim I_1(z) \sim \frac{e^z}{\sqrt{(2\pi Qt)}}.$$

Hence (6.10) is approximated by

$$\chi(x, t) = \frac{Q}{v} \exp\left(-\frac{x^2Q}{2v^2t} \right) \frac{1}{\sqrt{(2\pi Qt)}},$$

which agrees with (6.9), since $Q = 1/\tau$.

Now (6.9) has been derived on the assumptions that $N = t/\tau$ is large and $m^2 = O(N)$, i.e. $x^2/v^2 = O(t\tau)$, which are consistent with (6.11). Thus a first order agreement between (6.8) and (6.10) holds for t finite and $x/v = O(\sqrt{\tau})$, where τ is a small number. The requirement of Q being asymptotically equal to $1/\tau$ is not arbitrary; it is in fact necessary for the asymptotic equality of the rate of increase in variance per unit time for the two types of diffusion processes. For normal diffusion (6.9) this rate is $v^2\tau$, while for the continuous diffusion governed by the transition matrix (1.1) with $p = 1 - Q\tau$ the expression (4.9) shows that this rate is v^2/Q. Thus with a large rate of transition Q, the continuous diffusion becomes effectively the normal diffusion for $x/vt = O(1/\sqrt{Qt})$, i.e., for distances small compared with the maximum possible displacement.

7. DISTRIBUTIONS IN THE PRESENCE OF BARRIERS

General remarks. In this part the method of images has been employed to derive in a few cases the probabilities $\alpha(n, \nu)$ and $\beta(n, \nu)$ as modified by the presence of one or two barriers. The results are generalizations of those given by Chandrasekhar (1943) in as much as (i) they refer to the individual probabilities α and β and not to their sum, (ii) the initial probability of a particle moving to the right is c_1, not necessarily $\frac{1}{2}$ and (iii) they hold, except for absorbing barriers, for p not necessarily equal to q. It is to be noted that these formulae are applicable, only when the possible velocities u_1, u_2 are equal in magnitude. For it has been tacitly implied in the reasoning employed that in a path between two consecutive contacts with a barrier, the partial paths to the right are as many in number as those to the left, so that every such path has an image path between those very contacts. This is possible only if $u_1 = -u_2 = u$.

Reflecting barriers. First suppose that there is a reflecting barrier at $x = hd$, where $d = u\tau$ is the distance moved in each step. Denote by $\alpha^{(r)}(n, \nu)$ and $\beta^{(r)}(n, \nu)$ the probabilities $\alpha(n, \nu)$ and $\beta(n, \nu)$ as modified by the presence of the reflecting barrier, with $\nu \leqslant h$ and $h > 0$. Then we show that[1]

$$\alpha^{(r)}(n, \nu) = \alpha(n, \nu)+\beta(n, 2h-\nu),\ \nu \leqslant h; \qquad \dots \quad (7.1)$$

$$\beta^{(r)}(n, \nu) = \beta(n, \nu)+\alpha(n, 2h-\nu),\ \nu < h. \qquad \dots \quad (7.2)$$

In Figure 1 below is defined a 'family' of paths to comprise such paths as coincide when the parts of them to the right of the barrier are replaced by their images on the left; thus the paths FADP, FADP′, FA′DP and FA′DP′ constitute a family.

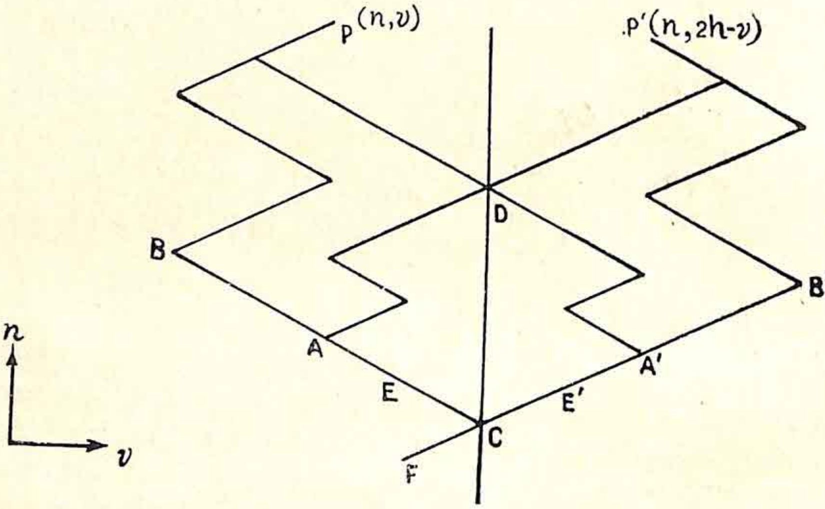


Fig. 1. 'Family' of paths.

It is first proved that if μ is the probability of a path which contributes to $\alpha^{(r)}(n, \nu)$, μ_1 the sum of the probabilities of those paths of the same family which contribute to $\alpha(n, \nu)$ and μ_2 the corresponding sum for $\beta(n, 2h-\nu)$, then

$$\mu = \mu_1+\mu_2. \qquad \dots \quad (7.3)$$

Consider a path FCBP, say, having single contact with the barrier; here P is the point (n, ν) and P′ its image $(n, 2h-\nu)$. This path contributes to $\alpha^{(r)}(n, \nu)$. If in the presence of the barrier its probability is μ, then in the absence of the barrier the probability attached to the path is μq, because the reflection at C is a certainty in the presence of the barrier but has a probability q in its absence. The probability μ_2 of the other member path of this family, viz., FCB′P′ is μp, since except for a continuation of direction at C the reversals and continuations in this path correspond to those of FCBP. Thus $\mu_1 = \mu q$ and $\mu_2 = \mu p$, which verifies (7.3).

---

[1] The formula (7.2) for $\beta^{(r)}(n,\nu)$ can be written from (7.1) by interchanging α and β and admits of a similar proof. This remark applies to the following sections as well, but the range of validity, as a little thought would show, is $\nu < h$.

In case there are two contacts at C and D, say, and μ is the probability of the path FCADP in the presence of the barrier, then this path and its homologues FCA'DP, FCADP', and FCA'DP' in the absence of the barrier have respectively the probabilities μq^2, μp^2, μqp and μpq. Thus

$$\mu_1 = \mu q^2+\mu p^2 \text{ and } \mu_2 = \mu qp+\mu pq, \text{ so that } \mu_1+\mu_2 = \mu(p+q)^2 = \mu,$$

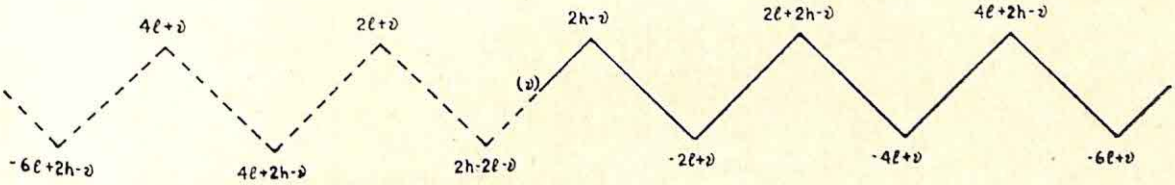
which verifies (7.3). The argument is similarly extended, whatever be the number of contacts.

The paths contributing to $\alpha^{(r)}(n, \nu)$ but not touching the barrier contribute also to $\alpha(n, \nu)$ the same amount; they have no homologues reaching beyond the barrier and thus (7.3) is again satisfied.

Adding all the equations (7.3) for all possible paths contributing to $\alpha^{(r)}(n, \nu)$, we prove (7.1); (7.2) is similarly proved.

Two reflecting barriers. When there are two reflecting barriers at $x/d = h$ and at $x/d = -k$, say, then we have to consider the reflections of the images themselves. The sequences of the successive images are shown in the following scheme :

Images in the barrier h



Images in the barrier $-k = h-l$, say.

Fig. 2.

In this scheme $2h-\nu$ is the position of the image of ν in $x/d = h$; $-2l+\nu$ is the position of the image of the former image in $x/d = h-l$, and so on. On the left, $2h-2l-\nu$ is the position of the image of ν in $x/d = -k$; $2l+\nu$ is the image of this in $x/d = h$, etc. Noting that after every reflection the path changes over its contribution from $\alpha(n, \nu)$ to $\beta(n, \nu)$ and vice versa, we have

$$\alpha^{(r)}(n, \nu) = \sum_{j=-\infty}^{\infty} \{\alpha(n, \nu+2jl)+\beta(n, -\nu+2h-2jl)\}. \qquad \dots \quad (7.4)$$

For a fixed n, only a finite number of terms will be non-vanishing in (7.4), for

(i) if j is so large that $|\nu+2jlJ > n$, then $\alpha(n, \nu+2jl) = 0$, and (*ii*) if j is so large that $|-\nu+2h-2jl| > n$, then

$$\beta(n, -\nu+2h-2jl) = 0.$$

For $k = \infty$, (7.4) reduces to (7.1).

A partly reflecting barrier. We first prove a lemma.

Lemma. *Let $\alpha_m(n, \nu)$ denote the probability that a particle at the end of the n-th step is reaching ν from the left after having been at $x/d = h$ just m times and similarly for $\beta_m(n, \nu)$; then*

$$\alpha_m(n, \nu)/\beta_m(n, 2h-\nu) = \{1+(q-p)^m\}/\{1-(q-p)^m\}. \qquad \ldots \quad (7.5)$$

Note that no barrier is assumed to be present.

Proof: For a given sequence of points $C_r\,(n_r, h)$, $r = 1, \ldots, m$, with a common abscissa[1] h and increasing ordinates, let L_m (or R_m) = probability of leaving the ordinate at h with the velocity $-\,u$ (or u) after the m-th contact at C_m.

Now an even number (including zero) of crossings of the ordinate occurs in a path contributing to L_m and an odd number in one contributing to R_m. Consequently from a family of homologous paths having contact with the ordinate at the m given points, each path contributing to L_m has a term corresponding to it in the expansion of $(p+q)^m$ with an even power of p, while each one of the paths contributing to R_m is corresponded by a term with an odd power in the expansion. Hence

$$\frac{L_m}{R_m} = \frac{(p+q)^m+(q-p)^m}{(p+q)^m-(q-p)^m} = \frac{1+(q-p)^m}{1-(q-p)^m}.$$

Now to every path L_m leading to $\alpha(n, \nu)$ there is an image path contributing to $\beta(n, 2h-\nu)$. Hence if $\alpha'_m(n, \nu)$ and $\beta'_m(n, 2h-\nu)$ denote respectively $\alpha(n, \nu)$ and $\beta(n, 2h-\nu)$ when the paths touch or cross the ordinate h at the m given points, we have

$$\frac{\alpha'_m(n, \nu)}{\beta'_m(n, 2h-\nu)} = \frac{L_m}{R_m} = \frac{1+(q-p)^m}{1-(q-p)^m}.$$

Using Σ to denote summation extending over all possible sequences of the m points of contact, we have

$$\frac{\alpha_m(n, \nu)}{\beta_m(n, 2h-\nu)} = \frac{\Sigma\alpha'_m(n, \nu)}{\Sigma\beta'_m(n, 2h-\nu)} = \frac{1+(q-p)^m}{1-(q-p)^m},$$

and the lemma is proved. This can be used to investigate the distribution in the presence of a partly reflecting barrier at $x = hd$, at which the probability of reflection is $q-p$ (> 0) and that of absorption is $2p$. Denote by $\alpha^{(p)}(n, \nu)$ and $\beta^{(p)}(n, \nu)$ the probabilities $\alpha(n, \nu)$ and $\beta(n, \nu)$ as modified by the presence of the partly reflecting barrier ; it is now shown that

$$\alpha^{(p)}(n, \nu) = \alpha(n, \nu)-\beta(n, 2h-\nu), \quad \nu \leqslant h. \qquad \ldots \quad (7.6)$$

---

[1] In the pair (n, ν), n is the ordinate and ν the abscissa.

Since the probability of reflection is $q-p$, we have, using the suffix m as in the lemma,

$$\alpha_m^{(p)}(n, \nu) = (q-p)^m \text{ times the probability of arriving at } (n, \nu) \text{ from the left in the presence of a perfectly reflecting barrier}$$

$$= (q-p)^m\{\alpha_m(n, \nu)+\beta_m(n, 2h-\nu)\}, \text{ by } (7.1),$$

$$= \alpha_m(n, \nu)-\beta_m(n, 2h-\nu),$$

on substituting for $(q-p)^m$ from (7.5). Summing now for all possible values of m, we are led to (7.6).

Absorbing barrier and $p = q$. In the general case, q not necessarily equal to p, although the simple relation (7.5) exists, it is difficult to determine either $\alpha_m(n, \nu)$ or $\beta_m(n, 2h-\nu)$ explicitly, for the ratio of the probabilities of the corresponding paths that contribute to $\alpha(n, \nu)$ and to $\beta(n, 2h-\nu)$ is not always the same, being $p : q$ for some and $q : p$ for other. But when $p = q$ the corresponding paths have equal probabilities, and using the superscript (a) for the presence of the absorbing barrier, we have

$$\alpha^{(a)}(n, \nu) = \alpha(n, \nu)-\beta(n, 2h-\nu), \quad \nu \leqslant h. \qquad \ldots \quad (7.7)$$

This is the relation (7.6), since when $p = q$, the probability of reflection becomes zero and the partly reflecting barrier becomes an absorbing one.

Alternatively this formula may be derived from the reasoning that to every path leading to (n, ν) from the left after having been at least once at $x = hd$, there corresponds a path of the same probability that leads to $(n, 2h-\nu)$ from the right. The right member of (7.7), therefore, represents the sum of the probabilities of all those paths which lead to (n, ν) from the left without having been at $x = hd$. Thus (7.7) is established.

Formula (7.7) gives for $\nu = h$ the probability of a particle being at the barrier for the first time; this is $\alpha(n, h)-\beta(n, h)$. Adding $\beta(n, h)$ to both the terms, we have

$$\alpha^{(a)}(n, h) = \gamma(n, h)-2\beta(n, h) = \gamma(n, h)-\gamma(n-1, h+1), \quad \text{[c.f. Chandrasekhar, p. 6]}$$

for in order to be able to contribute to $\beta(n, h)$ a particle must be at $(n-1, h+1)$ and from that position it is equally likely to assume the coordinates (n, h) or $(n, h+2)$.

Two barriers, at h *and* $-k$. The formulae of the preceding sections are easily extended to the case of two barriers. Writing for brevity $\alpha(n, \nu) \equiv \alpha_\nu$, etc., and employing an obvious notation and the scheme of images given in § 7, we have

$$\alpha^{(pp)}(n, \nu) = \alpha_\nu-\beta_{2h-\nu}+\alpha_{-2l+\nu}-\beta_{2l+2h-\nu}+\alpha_{-4l+\nu}-\cdots-\beta_{2h-2l-\nu}+\alpha_{2l+\nu}-\beta_{-4l+2h-\nu}+\alpha_{4l+\nu}-\cdots$$

$$= \sum_{m=-\infty}^{\infty} (\alpha_{\nu+2ml}-\beta_{2h-\nu+2ml}).$$

Taking $p = q = \frac{1}{2}$ we have formulae for two absorbing barriers.

Lastly, when there is a partly reflecting barrier at h and a reflecting one at $-k$, the scheme of images and (7.6) give

$$\alpha^{(rp)}(n, \nu) = \sum_{m=-\infty}^{\infty} (-)^m(\alpha_{\nu-2ml}-\beta_{2h-\nu-2ml}).$$

For $p = \frac{1}{2}$ this refers to an absorbing barrier at h and a reflecting one at $-k$.

I am grateful to Professor M. S. Bartlett and Mr. J. E. Moyal for their kind interest in this investigation.

References

Bartlett, M. S. (1955): *An Introduction to Stochastic Processes*, Cambridge University Press, London.

Chandrasekhar, S. (1943): Stochastic problems in physics and astronomy. *Rev. Mod. Phys.*, **15**.

Feller, W. (1953): *An Introduction to Probability Theory and its Applications*, I, John Wiley & Sons, New York.

Goldstein, S. (1951): On diffusion by discontinuous movements and on the telegraph equation. *Quart. Jour. Mech. and Applied Math.*, **4**.

Taylor, G. I. (1921-22): Diffusion by discontinuous movements. *Proc. London Math. Soc.*, **22**.

Paper received : April, 1956.

Revised : July, 1957.

ON THE DISTRIBUTION OF THE MAXIMUM VALUE OF AN EQUALLY CORRELATED SAMPLE FROM A NORMAL POPULATION

By A. KUDÔ

Kyushu University, Fukuoka, Japan
and
Indian Statistical Institute, Calcutta

SUMMARY. The density and the distribution function of equivariance multivariate normal sample are derived. The distribution function of the maximum of sample is also derived and the table for it is given.

1. Introduction

The purpose of this paper is to give the distribution of the maximum of an equally correlated sample from a normal universe. As an application we shall give the distribution of the statistic for the testing of outlying observation. The table necessary for the calculation of the distribution of the maximum value has been calculated on a small scale in the Indian Statistical Institute. The scale of the table was decided for the sake of another application, confidence interval for the extreme value in future observations. The confidence interval for it is discussed in Kudô (1954).

2. The derivation of the distribution

Let $(x_1, x_2, ..., x_N)$ be distributed normally with a common mean value zero and a common variance σ^2, and we assume that they are equally correlated with correlation coefficient ρ. As the correlation matrix should be positive, we should have $\rho \geqslant 1/N-1$. As the variance-covariance matrix is

$$\Lambda = \sigma^2 \begin{pmatrix} 1 & & \rho \\ & 1 & \\ & \ddots & \\ \rho & & 1 \end{pmatrix} = (1-\rho)\sigma^2 \begin{pmatrix} 1 & & 0 \\ & 1 & \\ & \ddots & \\ 0 & & 1 \end{pmatrix} + \rho\sigma^2 \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & 1 \\ \vdots & \vdots & & \vdots \\ 1 & 1 & \dots & 1 \end{pmatrix},$$

we have the characteristic function

$$\varphi(\underset{\sim}{t}) = \exp(-\tfrac{1}{2}\underset{\sim}{t}'\Delta\underset{\sim}{t}) = \exp\left(-\tfrac{1}{2}(1-\rho)\sigma^2\sum_{i=1}^{n}{}^2 - \tfrac{1}{2}\rho\sigma^2\left(\sum_{i=1}^{n} t_i\right)^2\right)$$

$$= \sum_{K=0}^{\infty} \frac{1}{K!}\left(\frac{-\rho\sigma^2}{2}\right)^K \left(\sum_{i=1}^{N} t_i\right)^{2k} \exp\left[-\tfrac{1}{2}(1-\rho)\sigma^2 \sum_{i=1}^{N} t_i^2\right]$$

$$= \sum_{K=0}^{\infty} \frac{(2K)!}{k!}\left(\frac{-\rho\sigma^2}{2}\right)^k \sum_{n_1+n_2+..+n_N=2k} \frac{t_1^{n_1}}{n_1!}\frac{t_2^{n_2}}{n_2!}\cdots\frac{t_N^{n_N}}{n_N} \exp\left[-\tfrac{1}{2}(1-\rho)\sigma^2 \sum_{i=1}^{N} t_i^2\right].$$

As this series is dominated by an exponential function, we can take the inverse Fourier transformation term by term.

Let $f(x)$ be an integrable function, $\varphi(t)$ be its Fourier transform, and $f(x)$ be differentiable up to n-th order. Then we can easily verify that the Fourier transform of the n-th derivative of $f(x)$ is $(-it)^n\varphi(t)$. Hence the density function of $(x_1, x_2, \ldots, x_N)$ is given by

$$f(x_1, x_2, \ldots, x_N) = \sum_{K=0}^{\infty} \frac{(2K)!}{K!} \left(\frac{\rho\sigma^2}{2} \right)^K (\sqrt{1-\rho}\,\sigma)^{2k+N} \sum_{n_1+\ldots+n_N=2k} \frac{1}{n_1!} F^{(n_1)}\left(\frac{x_1}{\sqrt{1-\rho}\,\sigma} \right) \cdots$$

$$\cdots \frac{1}{n_N!} F^{(n_N)} \left(\frac{x_N}{\sqrt{1-\rho}\,\sigma} \right),$$

where $F^{(n)}(x)$ is the n-th derivative of the standardized normal distribution function. Therefore the distribution function of $(x_1, x_2, \ldots, x_N)$ is

$$F(x_1, x_2, \ldots, x_N) = \sum_{K=0}^{\infty} \frac{(2K)!}{K!} \left(\frac{\rho\sigma^2}{2} \right)^K \{(1-\rho)\sigma^2\}^{k+N} \sum_{n_1+\ldots+n_N=2K} \frac{1}{n_1!} F^{(n_1)} \left(\frac{x_1}{\sqrt{1-\rho}\,\sigma} \right) \cdots$$

$$\cdots \frac{1}{n_N!} F^{(n_N)} \left(\frac{x_N}{\sqrt{1-\rho}\,\sigma} \right).$$

Let $a_n\ (x) = \frac{1}{n!} F^{(n)}\ (x)$ and then we have

$$F(x_1, x_2 \ldots, x_N) = \sum_{K=0}^{\infty} \frac{(2K)!}{K!} \left(\frac{\rho\sigma^2}{2} \right)^K ((1-\rho)\sigma^2)^{k+n} \sum_{n_1+n_2+\ldots n_N=2K} a_{n_1} \left(\frac{x_1}{\sqrt{1-\rho}\ \sigma} \right) \cdots$$

$$\cdots a_{n_N} \left(\frac{x_N}{\sqrt{1-\rho}\ \sigma} \right).$$

In case $x_1 = x_2 = \ldots = x_N = x$, it has a special expression :

$$F(x, \ldots, x) = Pr\ (\max x_i < x) = \sum_{K=0}^{\infty} \frac{(2K)!}{K!} \left(\frac{\rho\sigma^2}{2} \right)^K \{(1-\rho)\sigma^2\}^{k+n} A_{n,2k} \left(\frac{x}{\sqrt{1-\rho}\ \sigma} \right),$$

where $A_{n,K}\ (x)$ are defined in recurrent ways,

$$A_{1,k}\ (x) = a_k\ (x), \quad A_{n,K}\ (x) = \sum_{i=0}^{K} a_i\ (x)\ A_{n-1,k-i}\ (X).$$

3. TABLES

The tables were calculated in the following way. At first we tabulated $a_n\ (x)$ up to the seventh decimal place. Then $A_{n,K}(x)$ were tabulated by I.B.M. machines. The relations

$$\sum_{K=0}^{8} A_{n,K}(x) = \sum_{K=0}^{8} A_{n-1,K}\ (x) \sum_{i=0}^{8-k} a_i\ (X)$$

were used for check purpose at each stage of calculation and the numbers were listed upto the fifth decimal place. To examine the accuracy, another tabulation was carried out for $x = 3$ with ten decimal places, and there was found no discrepancy.

TABLES OF $A_{N,n}(X)$

x/n	0	2	4	6	8
			$N = 1$		
1.8	0.96407	−.07106	−.00142	.00136	−.00017
1.9	.97128	−.06234	−.00317	.00140	−.00015
2.0	.97725	−.05399	−.00450	.00135	−.00012
2.1	.98214	−.04618	−.00543	.00124	−.00008
2.2	.98610	−.03902	−.00598	.00108	−.00005
2.3	.98928	−.03258	−.00622	.00090	−.00002
2.4	.99180	−.02687	−.00618	.00070	.00001
2.5	.99379	−.02191	−.00593	.00051	.00003
2.6	.99534	−.01766	−.00553	.00034	.00004
2.7	.99653	−.01407	−.00503	.00019	.00005
2.8	.99744	−.01108	−.00447	.00006	.00005
2.9	.99813	−.00863	−.00389	−.00004	.00005
3.0	.99865	−.00665	−.00332	−.00011	.00004
3.1	.99903	−.00506	−.00279	−.00016	.00004
3.2	.99931	−.00381	−.00230	−.00019	.00003
3.3	.99952	−.00284	−.00187	−.00019	.00002
3.4	.99966	−.00209	−.00149	−.00019	.00002
3.5	.99977	−.00153	−.00118	−.00018	.00001
			$N = 2$		
1.8	0.92943	−.13077	.00696	.00308	−.00074
1.9	.94339	−.11678	.00148	.00352	−.00063
2.0	.95502	−.10261	−.00296	.00361	−.00049
2.1	.96459	−.08878	−.00633	.00343	−.00034
2.2	.97239	−.07570	−.00867	.00306	−.00020
2.3	.97867	−.06365	−.01009	.00258	−.00008
2.4	.98367	−.05280	−.01074	.00206	.00001
2.5	.98762	−.04324	−.01078	.00154	.00007
2.6	.99070	−.03497	−.01035	.00106	.00012
2.7	.99308	−.02793	−.00960	.00065	.00013
2.8	.99490	−.02204	−.00865	.00032	.00014
2.9	.99627	−.01719	−.00761	.00006	.00013
3.0	.99730	−.01326	−.00654	−.00013	.00011
3.1	.99807	−.01011	−.00552	−.00026	.00009
3.2	.99863	−.00762	−.00457	−.00033	.00007
3.3	.99903	−.00568	−.00372	−.00037	.00005
3.4	.99933	−.00419	−.00298	−.00037	.00004
3.5	.99953	−.00305	−.00235	−.00035	.00002

TABLES OF $A_{N,n}$ (X) *(continued)*

$x\backslash n$	0	2	4	6	8
			$N = 3$		
1.8	0.89604	−.18010	.02277	.00373	−.00172
1.9	.91630	−.16387	.01246	.00532	−.00154
2.0	.93329	−.14614	.00373	.00605	−.00124
2.1	.94736	−.12794	−.00321	.00609	−.00088
2.2	.95887	−.11011	−.00833	.00563	−.00054
2.3	.96817	−.09326	−.01178	.00486	−.00026
2.4	.97561	−.07781	−.01376	.00395	−.00003
2.5	.98149	−.06400	−.01457	.00302	.00011
2.6	.98608	−.05193	−.01447	.00214	.00021
2.7	.98964	−.04159	−.01372	.00139	.00025
2.8	.99235	−.03289	−.01255	.00077	.00025
2.9	.99441	−.02569	−.01115	.00028	.00023
3.0	.99596	−.01983	−.00966	−.00007	.00020
3.1	.99710	−.01513	−.00818	−.00030	.00017
3.2	.99794	−.01141	−.00680	−.00045	.00013
3.3	.99855	−.00851	−.00555	−.00052	.00010
3.4	.99899	−.00628	−.00445	−.00054	.00007
3.5	.99930	−.00458	−.00351	−.00052	.00004
			$N = 4$		
1.8	.86384	−.21991	.04393	.00235	−.00294
1.9	.88999	−.20410	.02847	.00599	−.00285
2.0	.91206	−.18485	.01477	.00806	−.00241
2.1	.93043	−.16381	.00346	.00879	−.00180
2.2	.94554	−.14235	−.00525	.00851	−.00117
2.3	.95779	−.12145	−.01142	.00757	−.00061
2.4	.96761	−.10191	−.01532	.00628	−.00018
2.5	.97539	−.08420	−.01735	.00489	.00011
2.6	.98149	−.06855	−.01790	.00355	.00029
2.7	.98620	−.05504	−.01739	.00237	.00037
2.8	.98982	−.04361	−.01616	.00139	.00039
2.9	.99256	−.03412	−.01451	.00064	.00036
3.0	.99461	−.02637	−.01267	.00008	.00031
3.1	.99614	−.02013	−.01079	−.00029	.00026
3.2	.99725	−.01519	−.00899	−.00053	.00020
3.3	.99807	−.01133	−.00736	−.00065	.00014
3.4	.99865	−.00836	−.00591	−.00069	.00010
3.5	.99907	−.00610	−.00467	−.00068	.00006

TABLES OF $A_{,n}(X)$ (*continued*)

x n	0	2	4	6	8
			$N = 5$		
1.8	.83280	−.25105	.06864	−.00168	−.00409
1.9	.86443	−.23794	.04832	.00496	−.00441
2.0	.89131	−.21901	.02943	.00917	−.00396
2.1	.91381	−.19652	.01323	.01117	−.00310
2.2	.93239	−.17242	.00033	.01145	−.00211
2.3	.94752	−.14824	−.00915	.01054	−.00120
2.4	.95968	−.12512	−.01549	.00896	−.00047
2.5	.96933	−.10384	−.01915	.00709	.00003
2.6	.97691	−.08484	−.02068	.00524	.00035
2.7	.98278	−.06830	−.02063	.00358	.00050
2.8	.98729	−.05422	−.01951	.00220	.00055
2.9	.99071	−.04248	−.01771	.00112	.00051
3.0	.99327	−.03286	−.01557	.00032	.00045
3.1	.99517	−.02511	−.01333	−.00023	.00036
3.2	.99657	−.01896	−.01116	−.00057	.00028
3.3	.99759	−.01415	−.00915	−.00076	.00020
3.4	.99832	−.01044	−.00736	−.00083	.00014
3.5	.99884	−.00762	−.00583	−.00083	.00009
			$N = 6$		
1.8	.80288	−.27428	.09536	−.00864	−.00481
1.9	.83961	−.26583	.07095	.00182	−.00600
2.0	.87103	−.24886	.04701	.00899	−.00580
2.1	.89749	−.22621	.02568	.01294	−.00477
2.2	.91943	−.20045	.00817	.01424	.00339
2.3	.93736	−.17367	−.00512	.01364	−.00205
2.4	.95181	−.14746	−.01434	.01188	−.00093
2.5	.96332	−.12294	−.02002	.00959	−.00014
2.6	.97236	−.10079	−.02282	.00720	.00037
2.7	.97938	−.08135	−.02345	.00501	.00063
2.8	.98477	−.06471	−.02257	.00317	.00071
2.9	.98886	−.05078	−.02074	.00171	.00068
3.0	.99193	−.03933	−.01838	.00064	.00059
3.1	.99421	−.03008	−.01582	−.00011	.00048
3.2	.99588	−.02272	−.01328	−.00058	.00037
3.3	.99710	−.01697	−.01092	−.00085	.00027
3.4	.99798	−.01253	−.00881	−.00096	.00018
3.5	.99861	−.00914	−.00698	−.00097	.00012

TABLES OF $A_{N,n}$ (X) *(continued)*

$x \backslash n$	0	2	4	6	8
			$N = 7$		
1.8	.77403	−.29033	.12275	−.01859	−.00472
1.9	.81550	−.28820	.09542	−.00366	−.00738
2.0	.85122	−.27463	.06690	.00724	−.00780
2.1	.88146	−.25302	.04043	.01384	−.00676
2.2	.90664	−.22651	.01802	.01668	−.00502
2.3	.92730	−.19778	.00055	.01672	−.00317
2.4	.94401	−.16894	−.01195	.01497	−.00159
2.5	.95733	−.14149	−.01998	.01231	−.00043
2.6	.96782	−.11640	−.02433	.00939	−.00033
2.7	.97598	−.09421	−.02585	.00664	−.00073
2.8	.98225	−.07509	−.02537	.00429	.00088
2.9	.98701	−.05901	−.02360	.00242	.00086
3.0	.99059	−.04575	−.02108	.00103	.00076
3.1	.99325	−.03502	−.01824	.00006	.00062
3.2	.99520	−.02647	−.01537	−.00056	.00047
3.3	.99662	−.01978	−.01267	−.00091	.00034
3.4	.99764	−.01460	−.01024	−.00108	.00024
3.5	.99837	−.01066	−.00812	−.00111	.00015
			$N = 8$		
1.8	.74622	−.29987	.14971	−.03137	−.00344
1.9	.79208	−.30546	.12091	−.01159	−.00826
2.0	.83185	−.29657	.08853	.00372	−.00980
2.1	.86571	−.27705	.05712	.01366	−.00900
2.2	.89404	−.25064	.02968	.01861	−.00697
2.3	.91736	−.22060	.00773	.01967	−.00460
2.4	.93627	−.18958	−.00838	.01815	−.00247
2.5	.95139	−.15952	−.01909	.01522	−.00086
2.6	.96331	−.13169	−.02524	.01179	.00022
2.7	.97260	−.10687	−.02785	.00846	.00081
2.8	.97974	−.08535	−.02791	.00557	.00104
2.9	.98517	−.06717	−.02629	.00325	.00105
3.0	.98925	−.05214	−.02368	.00151	.00093
3.1	.99229	−.03994	−.02061	.00029	.00076
3.2	.99452	−.03021	−.01743	−.00050	.00059
3.3	.99614	−.02258	−.01440	−.00096	.00043
3.4	.99731	−.01668	−.01166	−.00118	.00029
3.5	.99814	−.01218	−.00926	−.00124	.00019

TABLES OF $A_{N,n}(X)$ (*continued*)

$x \backslash n$	0	2	4	6	8
			$N = 9$		
1.8	.71941	−.30352	.17530	−.04669	.00068
1.9	.76933	−.31797	.14669	−.02195	−.00838
2.0	.81293	−.31488	.11138	−.00169	−.01161
2.1	.85024	−.29844	.07540	.01224	−.01140
2.2	.88161	−.27291	.04294	.01987	−.00921
2.3	.90752	−.24217	.01631	.02237	−.00631
2.4	.92860	−.20940	−.00370	.02133	−.00357
2.5	.94548	−.17702	−.01736	.01826	−.00144
2.6	.95882	−.14666	−.02556	.01436	.00002
2.7	.96923	−.11933	−.02945	.01045	.00085
2.8	.97724	−.09550	−.03019	.00698	.00120
2.9	.98333	−.07527	−.02882	.00418	.00125
3.0	.98792	−.05849	−.02618	.00206	.00112
3.1	.99133	.04484	−.02291	.00057	.00092
3.2	.99383	−.03394	−.01946	−.00041	.00071
3.3	.99566	−.02537	−.01612	−.00099	.00052
3.4	.99697	−.01875	−.01306	−.00127	.00035
3.5	.99791	−.01369	−.01039	−.00136	.00023
			$N = 10$		
1.8	.69356	−.30187	.19874	−.06415	.00381
1.9	.74724	−.32610	.17211	−.03464	−.00748
2.0	.79443	−.32979	.13499	−.00905	−.01304
2.1	.83505	−.31730	.09497	.00945	−.01385
2.2	.86935	−.29339	.05759	.02034	−.01171
2.3	.89779	−.26251	.02616	.02473	−.00832
2.4	.92098	−.22842	.00203	.02446	−.00492
2.5	.93961	−.19400	−.01483	.02141	−.00218
2.6	.95435	−.16131	−.02531	.01709	−.00027
2.7	.96587	−.13160	−.03065	.01259	.00084
2.8	.97474	−.10553	−.03221	.00853	.00134
2.9	.98150	−.08330	−.03119	.00521	.00144
3.0	.98658	−.06480	−.02859	.00268	.00132
3.1	.99037	−.04972	−.02516	.00089	.00109
3.2	.99315	−.03766	−.02145	−.00029	.00084
3.3	.99518	−.02817	−.01781	−.00099	.00061
3.4	.99664	−.02082	−.01446	−.00135	.00042
3.5	.99768	−.01521	−.01151	−.00148	.00027

4. AN APPLICATION

Let x_i $(i = 1, 2, \ldots, N_1+N_2)$ be independently and identically distributed as $N(m,\sigma^2)$, where σ is known and m is unknown to us. Let $x_M = \max_{i=1,\ldots,N_1} x_i$ and $\bar{x} = \sum_{i=1}^{N_1+N_2} x_i/N_1+N_2$. The statistic $(x_M-\bar{x})/\sigma$ have been proved to be optimum in a certain sense for the testing of an outlying observation, under such a circumstance as we know *a priori* there may be only one observation among $x_i(i = 1, 2, \ldots, N_1)$ which is from another normal population with a same variance and a greater mean value (Kudô, 1956).

In case $N_2 = 0$, the distribution was given by Mackey (1935), at first, and then Nair (1948) and Grubbs (1950) gave another form of it and tabulated the tables of this distribution.

As $y_i = x_i-\bar{x}$ $(i = 1, 2, \ldots N_1)$ are equally correlated we have

$$Pr\left(\frac{X_M-\bar{x}}{\sigma} < \lambda\right) = \sum_{K=0}^{\infty} \frac{(2k)!}{k!}\left(\frac{-1}{2(N_1+N_2)}\right)^k A_{N,2k}(\lambda).$$

The convergence is found to be quicker as λ increases.

The following is the comparison of our approximation with the values given by Grubbs for $N_2 = 0$, $\lambda = 1.8$, 2.5 and 3.5.

	1.8		2.5		3.5	
	Grubbs	approximation	Grubbs	approximation	Grubbs	approximation
$N = 3$	.95877	.95936	.99670	.99642	.99997	1.00000
$N = 4$	.92480	.92530	.99222	.99209	.99989	.99989
$N = 5$	.89037	.89066	.98703	.98696	.99977	.99978
$N = 6$	.85646	.85675	.98151	.98146	.99962	.99963
$N = 7$	.82341	.82364	.97580	.97576	.99945	.99945
$N = 8$	.79139	.79155	.96999	.96997	.99927	.99927
$N = 9$	.76046	.76058	.96412	.96411	.99908	.99908
$N = 10$	.73063	.73071	.95823	.95823	.99888	.99887

5. ACKNOWLEDGEMENTS

The author is much indebted to Professor P. C. Mahalanobis for providing him with a generous scholarship to work at the Indian Statistical Institute. He is grateful to Dr. C. R. Rao and Dr. D. Basu for their encouragement and advice in this work. Finally to all those who did the tabulation part of this paper, especially to Mr. D. Bose and Mr. A Halder, the author's gratitude is equally great.

REFERENCES

GRUBBS, F. E. (1950) : Sampling criteria for testing outlying observations. *Ann. Math. Stat.*, **21**, 27.

KUDÔ, A. (1956) : On the testing of outlying observations. *Sankhyā*, **17**, 67.

——— (1956a) : On the confidence interval for the extreme value of a second sample from a normal universe. Bull. Math. Stat., **6**, 51.

MACKEY, A. T. (1935) : The distribution of the difference between the extreme observation and the sample mean in samples of n from a normal universe. *Biometrika*, **27**, 466.

NAIR, K. R. (1948) : The distribution of the extreme deviate from the sample mean and its studentized form. *Biometrika*, **35**, 118.

Paper received : December, 1955.

HYPERGEOMETRIC EXPANSIONS FOR INCOMPLETE MOMENTS OF THE BIVARIATE NORMAL DISTRIBUTION

By A. R. KAMAT

Fergusson College, Poona, India

SUMMARY. Hypergeometric expansions are found for incomplete and absolute moments of the bivariate normal distribution. The use of incomplete moments in problems of truncation is indicated and a table of incomplete moments upto the weight four for different values of correlation coefficient is given.

1. We have defined and obtained expressions for incomplete and absolute moments of the bivariate normal distribution elsewhere (Kamat, 1953). Thus the incomplete moment of the bivariate distribution was defined as

$$[m, n] = \int_0^\infty \int_0^\infty x^m y^n p(x, y)\, dx\, dy, \qquad \dots \quad (1)$$

where $$p(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2+y^2-2\rho xy)\right],$$

and it is the contribution from the positive quadrant to the absolute moment defined as

$$(m, n) = \int_{-\infty}^\infty \int_{-\infty}^\infty |x|^m |y|^n\, p(x, y)\, dx\, dy. \qquad \dots \quad (2)$$

The expressions for (m, n) are useful in studying the distribution of random variates of the type $a|x_1|+b|x_2|$ where x_1, x_2 are correlated normal variates. It may also be noted that the expressions $[m, n]$ yield the moments of the truncated normal distribution $p(x, y)$ defined as above for $x > 0$, $y > 0$. For in this case moments are given by

$$\mu'_{m,n} = \epsilon(x^m y^n) = \frac{[m, n]}{[0, 0]}\,. \qquad \dots \quad (3)$$

2. In this note we first obtain a result which is interesting mathematically, viz., alternative expressions for $[m, n]$ and (m, n) in the form of hypergeometric series in ρ^2. Thus

$$[m, n] = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_0^\infty \int_0^\infty x^m y^n p(x, y)\, dx\, dy$$

$$= \pi^{-1} 2^{\frac{1}{2}(m+n)} (1-\rho^2)^{\frac{1}{2}(m+n+1)} \int_0^\infty\int x^m y^n \exp\,(-x^2-y^2+2\rho xy) dx\, dy$$

$$= \pi^{-1} 2^{\frac{1}{2}(m+n)} (1-\rho^2)^{\frac{1}{2}(m+n+1)} \int\int_0^\infty x^m y^n e^{-x^2-y^2} \Big(\sum_{\gamma=0}^{\infty} \frac{(2\rho xy)^r}{r\,!} \Big) dx\, dy$$

$$= \pi^{-1} 2^{\frac{1}{2}(m+n-4)} (1-\rho^2)^{\frac{1}{2}(m+n+1)} \Big\{ \sum_{r=0}^{\infty} \Gamma\Big(\frac{m+r+1}{2}\Big) \Gamma\Big(\frac{n+r+1}{2}\Big) \frac{(2\rho)^r}{\gamma\,!} \Big\}$$

$$= \pi^{-1} 2^{\frac{1}{2}(m+n-4)} (1-\rho^2)^{\frac{1}{2}(m+n+1)} \Big\{ \Gamma\Big(\frac{m+1}{2}\Big) \Gamma\Big(\frac{n+1}{2}\Big) F\Big(\frac{m+1}{2}, \frac{n+1}{2}, \tfrac{1}{2}; \rho^2\Big) +$$

$$+ 2\rho\, \Gamma\Big(\frac{m+2}{2}\Big) \Gamma\Big(\frac{n+2}{2}\Big) F\Big(\frac{m+2}{2}, \frac{n+2}{2}, \tfrac{3}{2}; \rho^2\Big) \Big\}$$

where $$F(\alpha, \beta, \gamma\,; x) = \sum_{r=0}^{\infty} \frac{\alpha(\alpha+1) \ldots (\alpha+r-1)\, \beta(\beta+1) \ldots (\beta+r-1) x^r}{\gamma\,!\, \gamma(\gamma+1) \ldots (\gamma+r-1)}.$$

Now since $$F(\alpha, \beta, \gamma\,; x) = (1-x)^{\gamma-\alpha-\beta} F(\gamma-\alpha, \gamma-\beta, \gamma\,; x)$$

we have

$$[m, n] = 2^{\frac{1}{2}(m+n-4)} \pi^{-1} \Big\{ \Gamma\Big(\frac{m+1}{2}\Big) \Gamma\Big(\frac{n+1}{2}\Big) F\Big(-\frac{m}{2}, -\frac{n}{2}, \tfrac{1}{2}\,; \rho^2\Big) +$$

$$+ 2\rho\, \Gamma\Big(\frac{m+2}{2}\Big) \Gamma\Big(\frac{n+2}{2}\Big) F\Big(-\frac{m-1}{2}, -\frac{n-1}{2}, \tfrac{3}{2}\,; \rho^2\Big\}. \quad \ldots \quad (4)$$

And summing up over the four quadrants we get

$$(m, n) = 2^{\frac{1}{2}(m+n)} \pi^{-1} \Gamma\Big(\frac{m+1}{2}\Big) \Gamma\Big(\frac{n+1}{2}\Big) F\Big(-\frac{m}{2}, -\frac{n}{2}, \tfrac{1}{2}\,; \rho^2\Big). \quad \ldots \quad (5)$$

It may be observed that for m (or n) even, the first part in (4) terminates while for m (or n) odd, the second part terminates. Hence we obtain a finite series for $[m, n]$ if one of m, n is even and the other is odd, and a finite series for (m, n) if at least one of them is even (c.f. Kamat, 1953, pp. 25, 26).

3. Since the incomplete moments are likely to prove useful in problems of truncation we end this note with tables of numerical values of $[m, n]$ for different values of ρ. When one follows the moments approach for studying a distribution one does not ordinarily go beyond the fourth moment. We have therefore given values of $[m, n]$ for $m+n \leqslant 4$

and for ρ positive and negative at an interval of 0.1. They are calculated from the expressions for $[m, n]$ given in Kamat (1953). It is to be noted that $[m, n]$ reduces to zero for $\rho = -1$ for all m and n, and is therefore not tabulated. Secondly values of [0, 0] which are given here are also available in Karl Pearson's Tables (1931). (It may be mentioned here that the entry in Pearson tables for [0, 0] corresponding to $\rho = 0.9$ is not correct.)

TABLE 1 : VALUES OF $[m, n]$ FOR $m+n \leqslant 4$ AND $\rho = 0.1\ (0.1)\ 1.0$

ρ	[0, 0]	[1, 0]	[2, 0]	[1, 1]	[3, 0]	[2, 1]	[4, 0]	[3, 1]	[2, 2]
0.1	0.265942	0.219418	0.281778	0.184951	0.458584	0.241360	0.876688	0.398080	0.318768
0.2	0.282047	0.239365	0.313255	0.212349	0.517029	0.287238	0.999585	0.487344	0.398174
0.3	0.298493	0.259312	0.344041	0.241372	0.573081	0.337106	1.115017	0.585956	0.488864
0.4	0.315495	0.279260	0.373842	0.272066	0.625541	0.390963	1.219549	0.693668	0.591494
0.5	0.333333	0.299207	0.402249	0.304499	0.673215	0.448810	1.310122	0.810122	0.706748
0.6	0.352416	0.319154	0.428811	0.338774	0.714905	0.510646	1.384217	0.934834	0.835339
0.7	0.373407	0.339101	0.452970	0.375045	0.749413	0.576472	1.440062	1.067169	0.978033
0.8	0.397583	0.359048	0.473978	0.413560	0.775544	0.646286	1.476937	1.206302	1.135674
0.9	0.428217	0.378995	0.490653	0.454769	0.792100	0.720091	1.495686	1.351126	1.309238
1.0	0.500000	0.398942	0.500000	0.500000	0.797885	0.797885	1.500000	1.500000	1.500000

TABLE 2 : VALUES OF $[m, n]$ FOR $m+n \leqslant 4$ AND $\rho = 0.0\ (-0.1)\ -0.9$

ρ	[0, 0]	[1, 0]	[2, 0]	[1, 1]	[3, 0]	[2, 1]	[4, 0]	[3, 1]	[2, 2]
0.0	0.250000	0.199471	0.250000	0.159155	0.398942	0.199471	0.750000	0.318310	0.250000
−0.1	0.234058	0.179524	0.218222	0.134951	0.339300	0.161572	0.623312	0.248081	0.191232
−0.2	0.217953	0.159577	0.186765	0.112349	0.280855	0.127662	0.500414	0.187345	0.141826
−0.3	0.201507	0.139630	0.155959	0.091372	0.224804	0.097741	0.384982	0.135956	0.101136
−0.4	0.184505	0.119683	0.126158	0.072066	0.172343	0.071810	0.280450	0.093669	0.068505
−0.5	0.166667	0.099736	0.097751	0.054499	0.124670	0.049868	0.189877	0.060123	0.043252
−0.6	0.147584	0.079788	0.071189	0.038774	0.082980	0.031915	0.115783	0.034834	0.024661
−0.7	0.126592	0.059841	0.047030	0.025045	0.048472	0.017952	0.059938	0.017169	0.011967
−0.8	0.102416	0.039894	0.026022	0.013560	0.022341	0.007979	0.023062	0.006302	0.004326
−0.9	0.071783	0.019947	0.009347	0.004769	0.005785	0.001995	0.004314	0.001127	0.000762

The author is thankful to the referee who suggested the construction of these tables and to Shri A. G. Deshmukh, research scholar in the Fergusson College for computational assistance.

References

Kamat, A. R. (1953): *Biometrika*, **40**, 20.

Pearson, K. (1931): *Tables for Statisticians and Biometricians*, **2**, 78.

Paper received : January, 1957.

INCOMPLETE MOMENTS OF THE TRIVARIATE NORMAL DISTRIBUTION

By. A. R. KAMAT

Fergusson College, Poona, India

SUMMARY. Expressions for incomplete moments of the trivariate normal distribution upto the weight four are given in terms of the correlation coefficients and their use in some problems of truncation is indicated.

We have defined and obtained expressions for incomplete and absolute moments of the bivariate normal distribution elsewhere (1953). In that paper we have also derived expressions for the absolute moments of the trivariate normal distribution, viz.,

$$[l, m, n] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |x|^l \, |y|^m \, |z|^n \, p(x, y, z)dx\, dy\, dz,$$

where $p(x, y, z)$ is the trivariate normal probability density function with $\epsilon(x) = \epsilon(y) = \epsilon(z) = 0$, and $\sigma(x) = \sigma(y) = \sigma(z) = 1$ and with correlation coefficients $\rho_{12}, \rho_{13}, \rho_{23}$. The incomplete moments are defined as

$$[l, m, n] = \int_0^{\infty}\int_0^{\infty}\int_0^{\infty} x^l \, y^m \, z^n \, p(x, y, z)dx\, dy\, dz,$$

and it is clear that $[l, m, n]$ is the contribution of the positive octant to the absolute moment (l, m, n). The expressions for the incomplete moments were *not* given in the paper mentioned above as they were not of immediate use there. It is now found however that these formulae have useful applications. Consider for instance the trivariate normal distribution defined above, but truncated to lie in the region $x > 0$, $y > 0$, $z > 0$. It is easy to see that the moment expressions for this distribution will be given by

$$\mu'_{l,m,n} = \epsilon(x^l \, y^m \, z^n) = \frac{[l, m, n]}{[0, 0, 0]} \, .$$

We are giving below formulae for $[l, m, n]$ upto and including $l+m+n = 4$, only. The expressions become quite formidable beyond $l+m+n = 4$. Since however one hardly goes beyond the fourth moment of a linear combination of x, y, z these formulae should be adequate for ordinary use. The method of derivation is outlined in the paper mentioned above. In the following formulae usual symbols have been used. For instance, $\omega =$ the determinant of the correlation matrix

$$= \begin{vmatrix} 1 & \rho_{12} & \rho_{13} \\ \rho_{12} & 1 & \rho_{23} \\ \rho_{13} & \rho_{23} & 1 \end{vmatrix}$$

and $\rho_{12\cdot 3} =$ partial correlation coefficient $= \dfrac{\rho_{12} - \rho_{13}\,\rho_{23}}{\sqrt{(1-\rho_{13}^2)(1-\rho_{23}^2)}}$.

List of $[l, m, n]$:

$$[0, 0, 0] = (4\pi)^{-1}\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}\right)$$

$$[1, 0, 0] = (2\pi)^{-\frac{3}{2}}\left\{\frac{\pi}{2}+\sin^{-1}\rho_{23\cdot1}+\rho_{12}\left(\frac{\pi}{2}+\sin^{-1}\rho_{23\cdot2}\right)+\rho_{13}\left(\frac{\pi}{2}+\sin^{-1}\rho_{12\cdot3}\right)\right\}$$

$$\begin{aligned}[2, 0, 0] = (4\pi)^{-1}\Big\{&\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}+\omega\,\rho_{23}(1-\rho_{23}^2)^{\frac{1}{2}}+\\ &+(2\rho_{12}\rho_{13}-\rho_{23})(1-\rho_{23}^2)^{\frac{1}{2}}+\rho_{13}(1-\rho_{13}^2)^{\frac{1}{2}}+\rho_{12}(1-\rho_{12}^2)^{\frac{1}{2}}\Big\}\end{aligned}$$

$$[1, 1, 0] = (4\pi)^{-1}\left\{\rho_{12}\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}\right)+(1-\rho_{12}^2)^{\frac{1}{2}}+\rho_{13}(1-\rho_{13}^2)^{\frac{1}{2}}+\rho_{23}(1-\rho_{23}^2)^{\frac{1}{2}}\right\}$$

$$\begin{aligned}[3, 0, 0] = (2\pi)^{-\frac{3}{2}}\Big\{&2\left(\frac{\pi}{2}+\sin^{-1}\rho_{23\cdot1}\right)+(1-\rho_{23}^2)^{-1}\,\omega\,(2\rho_{12}\rho_{13}-\rho_{12}^2\,\rho_{23}-\rho_{13}^2\,\rho_{23})+\\ &+\rho_{12}(3-\rho_{12}^2)\left(\frac{\pi}{2}+\sin^{-1}\rho_{13\cdot2}\right)+\rho_{13}(3-\rho_{13}^2)\left(\frac{\pi}{2}+\sin^{-1}\rho_{12\cdot3}\right)\Big\}\end{aligned}$$

$$\begin{aligned}[2, 1, 0] = (2\pi)^{-\frac{3}{2}}\Big\{&\rho_{13}+2\rho_{12}\left(\frac{\pi}{2}+\sin^{-1}\rho_{23\cdot1}\right)+(1+\rho_{12}^2)\left(\frac{\pi}{2}+\sin^{-1}\rho_{13\cdot2}\right)+\\ &+(\rho_{23}+2\rho_{12}\,\rho_{13}-\rho_{13}^2\rho_{23})\left(\frac{\pi}{2}+\sin^{-1}\rho_{12\cdot3}\right)\Big\}\end{aligned}$$

$$[1, 1, 1] = (2\pi)^{-\frac{3}{2}}\left\{\omega^{\frac{1}{2}}+\sum_{1,2,3}(\rho_{23}+\rho_{12}\,\rho_{13})\left(\frac{\pi}{2}+\sin^{-1}\rho_{23\cdot1}\right)\right\}$$

$$\begin{aligned}[4, 0, 0] = (4\pi)^{-1}\Big\{&3\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}\right)+\rho_{23}\,\omega^2\,(1-\rho_{23}^2)^{-\frac{3}{2}}+\\ &+2\rho_{23}\,\omega\,(2-\rho_{12}^2-\rho_{13}^2)(1-\rho_{23}^2)^{-\frac{1}{2}}+\rho_{12}(5-2\rho_{12}^2)(1-\rho_{12}^2)^{\frac{1}{2}}+\\ &+\rho_{13}(5-2\rho_{13}^2)(1-\rho_{13}^2)^{\frac{1}{2}}+(1-\rho_{23}^2)^{\frac{1}{2}}\Big(\rho_{23}(2\rho_{12}^2+2\rho_{13}^2-5)+4\rho_{12}\,\rho_{13}(3-\rho_{12}^2-\rho_{13}^2)\Big)\Big\}\end{aligned}$$

$$\begin{aligned}[3, 1, 0] = (4\pi)^{-1}\Big\{&3\rho_{12}\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}\right)+(2+\rho_{12}^2)(1-\rho_{12}^2)^{\frac{1}{2}}+\\ &+\omega\,(1-\rho_{23})^{-\frac{1}{2}}(\rho_{13}+\rho_{12}\,\rho_{23}-2\rho_{13}\,\rho_{23}^2)+(1-\rho_{23}^2)^{\frac{1}{2}}(2\rho_{23}+3\rho_{12}\rho_{13}-2\rho_{13}^2\rho_{23})+\\ &+2(1-\rho_{23}^2)^{\frac{1}{2}}(\rho_{13}+\rho_{12}\,\rho_{13}-2\rho_{12}\,\rho_{23}+2\rho_{12}^2\,\rho_{13}-\rho_{13}^2\,\rho_{23}+2\rho_{13}\,\rho_{23}^2-2\rho_{12}\,\rho_{23}\,\rho_{13}^2)\Big\}\end{aligned}$$

$$\begin{aligned}[2, 2, 0] = (4\pi)^{-1}\Big\{&(1+2\rho_{12}^2)\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{23}\right)+3\rho_{12}(1-\rho_{12}^2)^{\frac{1}{2}}+\\ &+(\rho_{23}+4\rho_{12}\,\rho_{13}-2\rho_{13}^2\,\rho_{23})(1-\rho_{23}^2)^{\frac{1}{2}}+(\rho_{13}+4\rho_{12}\,\rho_{23}-2\rho_{23}^2\rho_{13})(1-\rho_{13}^2)^{\frac{1}{2}}\Big\}\end{aligned}$$

$$\begin{aligned}[2, 1, 1] = (4\pi)^{-1}\Big\{&(\rho_{23}+2\rho_{12}\rho_{13})\left(\frac{\pi}{2}+\sin^{-1}\rho_{12}+\sin^{-1}\rho_{13}+\sin^{-1}\rho_{32}\right)+\\ &+(2\rho_{12}+\rho_{13}\,\rho_{23})(1-\rho_{13}^2)^{\frac{1}{2}}+(2\rho_{13}+\rho_{12}\rho_{23})(1-\rho_{12}^2)^{\frac{1}{2}}+\\ &+(1+\rho_{12}^2+\rho_{13}^2)(1-\rho_{23}^2)^{\frac{1}{2}}\Big\}.\end{aligned}$$

REFERENCE

KAMAT, A. R. (1953): *Biometrika*, **40**, 20.

Paper received : July, 1956.

A NOTE ABOUT A FAMILY OF PERKS' DISTRIBUTION

By JOSEPH TALACKO

Marquette University, Milwaukee, U.S.A.

SUMMARY. It is well-known in the classical theory of probability, that there exists some bell-shaped univariate densities, which differ from all known symmetrical distributions (Gauss, Student, Cauchy, etc.). This paper is concerned with a class of such distributions called "Perks' function." So far only two particular cases are usually discussed, the logistic and the hyperbolic secant distributions. Here, the characteristic function is introduced for the whole class, the properties, moments and meaning of parameters is discussed and the logistic and hyperbolic secant densities are shown as the borderline case. In spite of the great importance and usefulness of the Gaussian-normal distribution, these distributions have found wide applications in statistical practice.

1. Perks' system of distribution functions

The original formal Perks' function was written as

$$h(x) = \sum_{r=0}^{m} a_r c^{rx} / \sum_{r=0}^{n} b_r c^{rx} \qquad \text{... (1.1)}$$

where a_r, b_r and c are real parameters. In this general form, this function has not been studied extensively. The actual applications are for $m = 1$ and $n = 2$, with $c = e^{-1}$. For example, for $a_0 = 0$, $m \leqslant 1$ and $n \leqslant 2$, we have a class of asymmetrical distributions which may be written as

$$h_1(x) = \frac{A_1 e^{-x}}{b_0 + B_1 e^{-x} + B_2 e^{-2x}} \, . \qquad \text{... (1.2)}$$

Symmetrical forms are of greater interest. In particular, one family has been studied extensively. If in (1.2) $b_0 = B_2$ and $k = B_1$, we may write

$$h_2(x) = \frac{A_1}{b_0 e^{x} + k + b_0 e^{-x}} \, .$$

Written in the form of a probability differential, we have in the probabilistic notation

$$d\phi(x) = \varphi(x)dx = \frac{c\,dx}{e^{x} + k + e^{-x}} \, , \qquad \text{... (1.3)}$$

where $x = (\tau - \mu)/\alpha$ is the standardized variable; μ is the mean and α the standard, non-normalized parameter of this distribution. The c is a probability density constant and $k > -2$ a real parameter. A density exists for all real $k > -2$.

In the classical probability theory two particular cases are known, and applied in statistics (Paul Levy, 1953; Perks, 1932; Talacko, 1948).

(a) For $k = 0$, we have the p.d.f. and c.d.f. as

$$\varphi_1(x) = \frac{2}{\pi} \cdot \frac{1}{e^{x} + e^{-x}} = \frac{1}{\pi} \, \frac{1}{\cosh x} = \frac{1}{\pi} \operatorname{sech} x, \qquad \text{... (1.4)}$$

$$\phi_1(x) = \tfrac{1}{2} + \frac{1}{\pi} \text{ arc tanh } (\sinh x). \qquad \text{... (1.5)}$$

This function is known as a distribution of a hyperbolic secant or cosine.

(b) For $k = 2$, we get the p.d.f. and c.d.f. as

$$\varphi_2(x) = \frac{e^{-x}}{(1+e^{-x})^2} = \frac{1}{e^x+2+e^{-x}} = \tfrac{1}{4}\,\mathrm{sech}^2\,\frac{x}{2}, \qquad \ldots \quad (1.6)$$

$$\phi_2(x) = \frac{1}{1+e^{-x}} = \tfrac{1}{2}+\tfrac{1}{2}\tanh x. \qquad \ldots \quad (1.7)$$

This has been called the logistic distribution.

2. PROPERTIES OF PARTICULAR CASES AND COMPARISONS

Let us have one family of the distribution (1.3), for $0 \leqslant k \leqslant 2$

$$\varphi(x) = \frac{c}{e^x+k+e^{-x}}. \qquad \ldots \quad (2.1)$$

To show the properties, we will find the ch.f.

$$E(e^{ixv}) = C(x, v) = \int_{-\infty}^{\infty} e^{ivx}\, dF(x), \qquad \ldots \quad (2.2)$$

which in our case (2.1), with $k = 2 \cos \psi$, is given by

$$C(x, v) = c \int_{-\infty}^{\infty} \frac{e^{ixv}\, dx}{e^x+2\cos\psi+e^{-x}},\ \begin{matrix} 0 \leqslant \psi \leqslant \frac{\pi}{2}, \\ |v| < 1, \end{matrix} \qquad \ldots \quad (2.3)$$

$$C(x, v) = c \int_{-\infty}^{\infty} \frac{e^{x(1+iv)}}{(e^x+e^{i\psi})(e^x+e^{-i\psi})} \qquad \ldots \quad (2.4)$$

and the corresponding contour integral

$$I(\psi) = C(x, v) = c \int_C \frac{e^{z(1+iv)}}{(e^z+e^{i\psi})(e^z+e^{i\psi})}. \qquad \ldots \quad (2.5)$$

After evaluation and simplification, we get the (ch.f.) in elementary form

$$C(x, v) = \frac{c\,.\,\pi}{\sin\psi} \cdot \frac{\sinh(\psi v)}{\sinh(\pi v)} \qquad \ldots \quad (2.6)$$

as a function of ψ.

For

$$\lim_{v\to 0} C(x,v) = \lim_{v\to 0} \frac{c\pi}{\sin\psi} \cdot \frac{\sinh \psi v}{\sin \pi v} = 1, \qquad \ldots \quad (2.7)$$

$$c = \frac{\sin\psi}{\psi},$$

and finally

$$C(x, v) = \frac{\pi}{\psi} \frac{\sinh(\psi v)}{\sinh(\pi v)}. \qquad \ldots \quad (2.8)$$

From (2.8) we find ch.f.'s for all distributions (1.3) :

(a) For $-2 < k < 0$

$$C_a(x, v) = \frac{\pi}{\psi} \frac{\sinh (\psi v)}{\sinh \ (\pi v)}, \left(\frac{\pi}{2} \leqslant \psi \leqslant \pi \right). \qquad \ldots \quad (2.9)$$

(b) For $k > 2$, using $= i\theta$, we get

$$C_b(x, v) = \frac{\pi}{\theta} \cdot \frac{\sin \theta v}{\sinh \ \pi v} . \qquad \ldots \quad (2.10)$$

The ch.f. for the hyperbolic secant distribution (1.4) we get, as a particular case of (2.8) for $\psi \rightarrow \pi/2$,

$$C_1(x, v) = \text{sech} \ (\pi v/2). \qquad \ldots \quad (2.11)$$

The ch.f. of the logistic distribution (1.6) for $\psi \rightarrow 0$ is

$$C_2(x, v) = \frac{\pi v}{\sinh \pi v} = \pi v \ . \cosh \pi v. \qquad \ldots \quad (2.12)$$

The ch.f.'s for the random variable $\tau = (\mu + \alpha x)$ in (1.3) are

$$C_1(\tau, v) = e^{i\mu v} \, \text{sech} \ (\pi \alpha v)/2) \qquad \ldots \quad (2.13)$$

and

$$C_2(\tau, v) = e^{i\mu v} \, \pi \alpha v \ \text{cosech} \ (\pi \alpha v). \qquad \ldots \quad (2.14)$$

These may be compared with three classical distributions :

(a) Normal, where for $T = (x - \mu)/\delta_x$

$$C_N \, (x, v) = \exp \, (i \mu v - \delta_x^2 v^2/2), \qquad \ldots \quad (2.15)$$

(b) Double exponential

$$C_{DE}(x, v) = \frac{\exp(i\mu v),}{1 + \alpha^2 v^2}, \qquad \ldots \quad (2.16)$$

and

(c) Cauchy,

$$C_C(x \ v) = \exp \, (i \mu v - \alpha \, | \, v \, |). \qquad \ldots \quad (2.17)$$

The natural parameter for the normal distribution is δ_x, for all others is α as the basic unit interval. Of course, for the hyperbolic secant and the logistic it is possible to introduce also the random variable

$$T = \frac{\mu - x}{\sqrt{\mu x, 2}}$$

which we may call a "normalized" standard variable.

The "standard" moments for the hyperbolic secant are given by the Euler numbers

$${}_H\Lambda_{2r} = | E_{2r} |, \quad r = 1, 2, 3, \ldots,$$

and for the logistic by

$${}_L\Lambda_{2r} = 2.3(2^{2r-1} - 1) \, | B_{2r} |,$$

where B_{2r} are the Bernoulli numbers. This is interesting in comparison to the normal distribution, with

$$_{N}\Lambda_{2r} = \frac{(2r)!}{2^r r!},$$

and to the double exponential where

$$_{DE}\Lambda_{2r} = \frac{(2r)!}{2^r}.$$

For the Cauchy distribution, these moments do not exist.

In particular we have

standard moment	normal	logistic	hyperbolic secant	double exponential
Λ_4	3	4.2	5	6
Λ_6	15	39.86	61	90

4. THE PROBABILISTIC MEANING AND THE DIVISIBILITY

Let us take as the basic population the double exponential distribution with $\mu = 0$ and $\alpha = 1$, with d.f. and ch.f.

$$dF(x) = \tfrac{1}{2}\, e^{-|x|}\, dx, \qquad \ldots \quad (3.1)$$

$$C(x, v) = \frac{1}{1+v^2}.$$

Let the random variable X from this population be defined by

$$X = a_1x_1+a_2x_2+\ldots+a_rx_r+\ldots . \qquad \ldots \quad (3.2)$$

One may construct a set of sampling distributions from this population which includes the normal, hyperbolic secant, logistic and others, if $a_1, a_2, \ldots, a_r$, are specially chosen infinite sequences.

(a) Let $a_1 = a_2 = \ldots = a_r = \ldots = 1/\sqrt{2n}$. We have

$$\sqrt{2n}\, X = \sum_{x=1}^{\infty} x_r.$$

The characteristic function

$$C(X, v) = \prod_{r=1}^{\infty} C\left(x_r, \frac{v}{\sqrt{n}}\right) = \prod_{r=1}^{\infty} \frac{1}{(1+v_r^2/2n)} = \lim_{n\to\infty}\left(1+\frac{v^2}{2n}\right)^{-n} = e^{-v^2/2}. \qquad \ldots \quad (3.3)$$

We get the normal distribution, where $\sqrt{2n}$ is a most important number, normalizing the result.

(b) Let $a_r = 1/n.\ r = 1, 2, 3 \ldots$. Then

$$nX = \sum_{r=1}^{\infty} x_r$$

$$C(X, v) = \lim_{n\to\infty}\left(1+\frac{v^2}{n^2}\right)^{-n} = 1. \qquad \ldots \quad (3.4)$$

We get a single point-degenerate distribution.

(c) Let $a_r = 2/\pi(2r-1)$. Then we get

$$X = (2/\pi) \cdot (x_1+x_2/3+x_3/5+\ldots+x_r/2r-1+\ldots). \qquad \ldots (3.5)$$

The characteristic function of such a random variable is

$$C(X, v) = \prod_{r=1}^{\infty} C_r\left(x_r, \frac{v}{\pi(r+\frac{1}{2})}\right) = \prod_{r=1}^{\infty} \frac{1}{1+\left(\frac{2v}{\pi 2r-1}\right)^2} = \frac{1}{\cosh v} = \frac{2}{e^v+e^{-v}}. \qquad \ldots (3.6)$$

The inverse Fourier transformation gives the probability desnity

$$E(e^{-ixv}) = \varphi(x) = \int_{-\infty}^{\infty} \frac{2e^{-ixv}}{e^v+e^{-v}}\, dv = \tfrac{1}{2} \operatorname{sech}(\pi x/2). \qquad \ldots (3.7)$$

This is a classical example of an exception to the central limit theorem. This theorem requires that $\sum_{r=1}^{\infty} \delta_r^2 \to \infty$, for $r \to \infty$.

But here $\sum_{r=1}^{\infty} \delta_r^2 = 1$ because

$$\sum_{r=1}^{\infty} \delta_r^2 = \frac{8}{\pi^2}\left(1+\frac{1}{3^2}+\frac{1}{5^2}+\ldots\right),$$

where $(\ldots) = \pi^2/8$. In many applications, the sum of the variable is indeed bounded so that such restrictions are natural and have a physical correspondence in many observations.

(d) In (3.2) let $a_r = 1/r$.

Then
$$X = x_1+x_2/2+x_3/3+x_4/4+\ldots+x_r/r+\ldots,$$

and
$$C(X, v) = \prod_{r=1}^{\infty} C\left(x_r, \frac{v}{r}\right) = \prod_{r=1}^{\infty}\left(1+\frac{v^2}{r^2}\right)^{-1}$$

$$= \frac{2\pi v}{e^{\pi v}-e^{-\pi v}} = \pi v \cdot \operatorname{cosech} \pi v. \qquad \ldots (3.8)$$

The density function is defined by the relation

$$\varphi_2(x) = E\left(e^{-ixv}\right) = \frac{e^{-x}}{(1+e^{-x})^2} \qquad \ldots (3.9)$$

it is the logistic distribution.

(e) Similar'y, in the same way, it is possible to show the relation between the hyperbolic secant and the logistic distribution. From the theory of infinite processes we know that the infinite product

$$\prod_{r=0}^{\infty} \cos x/2^r = \sin x/x. \qquad \ldots (3.10)$$

If we take ix as the variable, this becomes

$$\prod_{r=0}^{\infty} \cosh x/2^r = \sinh x/x.$$

In the hyperbolic secant population, let the random variable once more be a countable set

$$X = a_1x_1+a_2x_2+\ldots+a_rx_r+\ldots$$

where $a_r = 1/2^r$ for $r = 1, 2, 3, \ldots$. Then

$$C(X, v) = C_1(v) \, . \, C_2(v) \ldots C_r(v) \ldots$$

and after substitutions

$$C(X, v) = \prod_{r=0}^{\infty} 1/\cosh \frac{v}{2^r} = \frac{v}{\sinh v}. \qquad \ldots \quad (3.11)$$

This is nothing other than the (c.f.) of the logistic , with the density

$$\varphi_2(x) = \frac{1}{e^x+2+e^{-x}}.$$

The relations between the hyperbolic secant and logistic are simply

$$\varphi_1(x) = \frac{2}{\pi}\,\frac{1}{e^x+e^{-x}} = \frac{1}{\pi} \text{ sech } x, \; C_1(v) = \text{sech } \frac{\pi v}{2},$$

$$\varphi_2(x) = \frac{1}{e^x+2+e^{-x}} = \frac{e^{-x}}{(1+e^{-x})^2} = 1/4 \text{ sech}^2 \frac{x}{2}, \; C_2(v) = \pi v \text{ cosech } \pi v.$$

If we take the variable as πx, $C_2(x\pi, v) = v$ cosech v.

Let $f_3(x) = cx/\sinh x$. This is a new distribution with ch.f.

$$C_3(x, v) = C_1(v) \, . \, C_1(v) = C \, . \, \text{sech}^2 \frac{\pi v}{2}. \qquad \ldots \quad (3.12)$$

We may write the convolution

$$f_3(x) = C \int_{-\infty}^{\infty} \text{sech } y \text{ sech } (x-y) dy.$$

The result :

$$\varphi_3(x) = 2x \, . \text{ cosech } x. \qquad \ldots \quad (3.13)$$

References

Levy, Paul (1953): *Random Function: General Theory with Special Reference to Laplacian Random Functions,* University of California Publication, in Statistics, **1**, 331.

Perks, Wilfred (1932): On some experiments in the graduation of mortality matrixes. *Journal of the Institute of Actuaries*, **58**, 12.

Talacko, Joseph (1948): Mathematical theory of growth with special regard to population problems. *Aktuarske Vedy*, **8**, 1.

Wolfenden, Hugh H. (1942): *Fundamental Principles of Mathematical Statistics*, Stechert-Hafner, New York.

Paper received : August, 1956.

RECENT EXPERIMENTS IN STATISTICAL SAMPLING IN THE INDIAN STATISTICAL INSTITUTE[1]

By P. C. MAHALANOBIS

Indian Statistical Institute, Calcutta

INTRODUCTION

I was naturally gratified to receive in March 1946 an invitation to present a paper before the Royal Statistical Society during my visit to London. I had collected some material for this purpose and had written certain portions of the paper when I was suddenly obliged to leave India at the end of April to attend a session of the United Nations Statistical Commission in New York. Owing to unforeseen difficulties and pressure of work in connection with the Royal Society and Empire Scientific Conferences in London in June and July, I had no time to finish the paper. It was not even possible to circulate the tables which had to be presented before the meeting in a mimeographed form. I greatly appreciate the kindness shown to me in relaxing the rule about submitting the written paper in advance, and secondly in arranging the present meeting during the closed session. The present paper contains the observations made at the meeting as well as certain portions of my original notes. I acknowledge with thanks the help I have received from my young colleague, Prasad Banerjee, in putting this paper in its final shape.

Restricted scope of the paper. This paper has been deliberately named as "Recent Experiments in Statistical Sampling in the Indian Statistical Institute" to make it clear that I had no desire to speak on sample surveys in general, or to make any attempt to give an account of the many interesting and significant developments in statistical sampling which have taken place in recent years outside India. A large number of papers have been published on the subject in statistical and scientific journals, and Dr. F. Yates, in the paper recently presented before the Royal Statistical Society, has given a comprehensive review. It is not necessary for me to touch the ground covered by others. The aim of the present paper is to draw attention to certain experiments in statistical sampling conducted during the period 1937- 45 by the Indian Statistical Institute, of which I am the Honorary Secretary.

The Indian Statistical Institute. It will be convenient if I give at this stage a brief account of the Indian Statistical Institute itself, as this would explain the set-up in which the work was done. The Institute was inaugurated on December 17th, 1931, and was formally registered on April 28th, 1932, as a non-profit-making scientific society. Work began on a small scale, and the total expenditure in the first year (1932-33) was about £40. Special courses in statistics were started in the second year, which later developed into regular post-graduate training classes. Research work was also begun, at first on a part-time basis, but later with the help of a whole-time nuclear staff. A scheme for conducting examinations and awarding certificates of proficiency for computers and statisticians had been prepared in 1935-36 which was brought into effect from 1938. In 1941 the Institute helped in establishing a post-graduate department of statistics in the Calcutta University offering whole-time courses for the degrees in Statistics of M.A. and M.Sc.; and the actual teaching work was done in the Institute itself by a practically joint Institute and University staff. *Ad hoc* enquiries on a small scale were being undertaken from 1935, but project work began to develop on a large scale from 1937 with the initiation of a scheme for improving the forecast of the area under jute in Bengal. The project side developed rapidly during the war, as the Institute had to undertake various statistical enquiries and surveys on behalf of Government Departments. In 1945-46 the total volume of employment was about 750 man-years, divided about equally between the statistical and field branches, and the total expenditure was of the order of £100,000. Excepting for a comparatively small research grant of about £4,000, practically the whole of the income was derived in the form of contract grants from Government Departments for specific enquiries and projects. It is interesting to observe that the machinery was practically the same which had developed independently in the U.S.A., where large schemes of statistical investigations relating to the war effort were being done in different Universities and scientific institutions with the help of contract grants from the Government.

---

[1] This paper originally appeared in the *Journal of the Royal Statistical Society*, volume 109, part 4, 1946 and is being reprinted with the permission of the Royal Statistical Society—Editor.

Conditions of work. Several points deserve to be emphasized. Each project or enquiry had definite practical objectives which were laid down by the client (usually a Government Department); the permissible margin of error of the results was broadly indicated; the enquiry had to be conducted in accordance with a definite time schedule so as to enable interim or final reports to be submitted by particular dates; the whole work had to be done within the limits of an all-inclusive contract grant, so that any excess of expenditure would mean a financial loss to the Institute. Finally, owing to difficulties created by the war situation in North-east India, the human agency and the organizational side of the project had to be very carefully planned, and usually the field survey had to be carried out under great difficulties. The research that had to be undertaken was naturally of an applied type. Nevertheless fascinating theoretical problems were continually arising, some of which were and are being tackled on fundamental lines. In preparing the design of each sample survey the three most important considerations thus were time, cost, and the human agency—all, of course, in relation to the most important factor of the permissible margin of error which was usually stipulated in advance.

Statistical engineering. In the present paper I have very much in mind those problems of organization which arise when a sample survey has to be carried out on a very large scale. The difference in quantity is so great that it brings about practically a change in quality. The manufacture on a commercial scale of a chemical gives rise to problems of an entirely different type from those which have to be solved in small-scale working in a laboratory, and the difference is so great that we give expression to it by calling large-scale production a matter of chemical engineering rather than of pure chemistry. In the same way large-scale sample surveys may be appropriately called statistical engineering. A good deal of the present paper is in fact concerned with statistical engineering rather than the pure theory of sampling.

General observations

The sample enquiries discussed in the present paper fall into two broad types. A great deal of work has been done on sample surveys on an extensive scale covering whole provinces (50,000 to 140,000 square miles in extent in the same season), in which the object has been to obtain reliable estimates of the acreage, rate of yield per acre, and total production of important food and fibre crops like rice, wheat, jute, etc.; and of economic or demographic factors relating to indebtedness, unemployment, destitution, paddy land, plough cattle, birth rate, death rate, etc., of rural families. For such work the technique used has always been the area or grid method, in which comparatively small sampling units (of area or size ranging from about two acres to one square mile) were located at random, and information was collected for each sample-unit by direct physical inspection or investigation. Estimates for districts (about 2,500 square miles or so in area on an average) or the province as a whole were then obtained directly by multiplying observed averages by appropriate multiplying factors (equal to the ratio of the total geographical area of the region to the total area covered in the sample surveys).

In a second type of work the enquiry was more localized, and related to cost and level of living, housing, consumption of food, clothes, etc.; preferences for particular types of commodities; reactions to radio programmes; public opinion on various subjects, etc.

Terminology and classification of sample surveys

Before proceeding further it will be convenient to explain the terminology and classification of sample surveys. In the area method the smallest physical element which can be separately surveyed is called a *quad*. Typical examples are a single plant for determining the yield of crop; a single family or household (having food from the same kitchen) for ascertaining the cost of living; or a single individual in anthropometric or blood-group surveys. The whole region or field to be surveyed may be then considered to consist of a very large number of such ultimate physical elements or *quads*.

Quad and configurational sampling. Sampling may then proceed in two different ways. The ultimate physical element or quad may be adopted as the sampling unit, or a group of such elements taken together may be used as the unit for sampling; these two types are called "quad" and "configurational" (or "grid" or "cluster") sampling. The grid, in the area method, consists of a suitable number of adjoining quads or ultimate physical units. For example, in crop surveys in Bengal it was found that a suitable sampling unit was a grid of square shape and size 2.25 acre. Each grid thus consists of a compact group of adjoining quads.

It must be noted, however, that the sample unit may also consist of quads separated from one another, but in a fixed spatial configuration—for example, 4 quads at the corner of a square of a given size. It is convenient therefore to distinguish between the "grid" and the "cluster," and restrict the former to a compact group of adjoining individual units, and the latter to other types of configurational sampling where the quads are not adjoining but are arranged in some particular pattern in one or two dimensions. Every n-th individual unit in a linear series would also in this sense be an example of cluster or configurational sampling. From the present point of view all sampling methods may then be characterized as being of either (1) of *quad*, (2) of *grid*, or (3) of *cluster* type.

Zonal and non-zonal. The important point to be noted is that all the sample-units have to be located at random. But this can be done in two ways—namely, distributing the sample units over the region under survey as a whole, or by first dividing the whole region into a suitable number of sub-divisions or zones, allocating a suitable number of sample units to each zone, and then locating such sample units, grids, or clusters at random within each zone separately. This gives rise to two types—namely, (1) without zoning or non-zonal, and (2) zonal sampling.

Multi-stage. So far methods have been considered in which the sample units are located at random in one single stage over each zone or over the whole area under survey. In many enquiries this method cannot be used because of its high cost. For example, in estimating the rate of yield per acre of crops it is usually necessary to proceed by stages. A number of zones are first selected at random; within each selected zone a number of villages are next selected at random; within each village so selected a number of fields are then selected at random; and within each field so selected one or more sample cuts are finally located at random. In this method the act of randomization is performed, not in one single stage, but in successive stages. The coverage of the process of randomization takes place repeatedly over smaller and smaller regions. This type of sampling is called multi-stage.

Replicated networks of sample units. Information for each zone (or for the entire field under survey) may be collected in one single network of sample units; or the information may be collected separately for two or more networks of sample units. Each such network would give an independent estimate, and differences between such estimates supply immediately measures of the over-all or effective margin of error. A characteristic feature of the Statistical Institute's work has been the use of such independent replicated networks of samples.

It would be noticed that quad or configurational, zonal or non-zonal, and replicated sampling can be used with either the uni-stage or the multi-stage methods. Thus quad sampling may be used at one or more stages, and grid or cluster at another stage. In the same way, the sampling method may be non-zonal at one or more stages and zonal at others. There may be one single network of samples at one or more stages or more than one network at other stages. The present system of classification of sampling methods is thus flexible and comprehensive, and makes it possible to describe any particular design of sample survey in a precise and non-ambiguous manner. It is suggested that the present terminology (with such modifications that may be considered desirable) should be adopted as a standard system for the classification of sample surveys.

LARGE-SCALE SURVEYS

When the work is done on an extensive scale covering a large geographical area it is almost inevitable that the sample units must be located at a considerable distance from one another. The investigators have therefore to spend a good deal of time in travelling from one sample unit to another. The total time (or the total cost) of field operations thus consists of two broad portions—namely, the actual time or cost required for enumeration or investigation of the sample units, and the time required for the journeys between different sample units. In such surveys it would be obviously uneconomical to collect information for a single quad in each locality. It is therefore usually arranged to use a grid or a cluster—that is, a group of quads for investigation in each locality visited by the investigator.

Cost function. Large-scale sample surveys are necessarily expensive, and considerations of cost are therefore of great importance.

As already noted, in surveys covering large geographical areas the cost of field operations depends on the time required for actual enumeration or investigation and the time required for journeys from one sample unit to another. Collecting information for grids of a large size would naturally take more time. If grids of a large size are used, then the total number which can be surveyed must be necessarily small. Grids of a large size would be thus more widely scattered, so that the time required for journeys would be larger for each journey on an average, but there would be fewer journeys to perform. Working with grids of a small size it would be possible to use a large number so that the average distance apart would be less. The time for journeys on an average would be less, but a large number of journeys would have to be undertaken. The total cost of operations thus depends on the size of the grids and their total number. Working within a fixed budget this means that once the size of the grids is settled, their total number must also be determined by the budget limit.

In enquiries which are geographically more localized, but which are conducted on a sufficiently large scale, the cost of operation depends on the way the field survey is organized—that is, on the nature of the human agency employed for the work, the methods adopted for collecting the primary data, arrangements for checking and supervision, etc. In every large-scale enquiry the cost is, therefore, determined by the particular plan or design of survey proposed to be adopted.

Variance function. Cost, however, is not the only consideration. The precision of the final results is also important. The variance of grids or sample units decreases as the size (*i.e.*, area) of the grid or sample unit is increased. In crop-survey work it became clear quite early that the decrease in variance with an increase in the size of the grid was appreciably slower than the normal rate of decrease (namely, inversely as the size of the sample unit). What particular size of grid would be most economical would therefore depend on how the variance changes with the size of the grid. This necessitates a study of the variance function which gives the relation between the variance and the size of the grid. More generally, each way of distributing the sample units—that is, each particular design of sampling—involves its own appropriate sampling error, so that corresponding to each design there exists its specific variance function.

Optimum solutions. It is now possible in terms of the cost and variance functions to state the conditions for an optimum or most economical design for sampling. The aim may be stated in either of the following two alternative forms: (*a*) to determine the size and distribution or density of grids or sample units in such a way that the variance (or margin of error) of the final estimate is a minimum when the total cost of the sample survey is fixed; or, alternatively, (*b*) to determine the size and distribution of sample units so as to reduce the variance (or margin of error) of the final estimate to any desired level at a minimum cost. It has been shown elsewhere how concrete solutions (which are the same for either of the alternative forms) can be obtained with the help of an empirical knowledge of the variance and cost functions.

Two points are worth noting. The above approach is indispensable when the scale of operations is sufficiently large, and individual grids or sampling units are so widely separated that an appreciable amount of time and cost is incurred in journeys between grids or sample units, or when a large number of investigators are employed for the field survey so that the cost depends materially on the way in which the field staff is organized. Secondly, such large-scale sample surveys usually give rise to many problems involving the human agency which have engaged the special attention of the Indian Statistical Institute for a long time, and to which more detailed references have been made in later sections.

Exploratory (or sequential) development

The planning of a sample survey thus has several aspects—namely, (1) zoning and/or stratification; (2) size of the grid or sample unit; (3) arrangements for replication; (4) density or distribution of grids in different zones or strata; (5) preparation of forms and schedules; (6) structure and organization of the field staff; (7) arrangements for the statistical processing of the material, etc. It has been explained above how the efficiency of the design of the sample survey can be maximized from joint considerations of variance and cost functions. To use such methods it is necessary to have previous information about those two functions. Other relevant information relating to the region or universe to be surveyed is also of great value in preparing the design. In fact, the greater the amount and accuracy of such information the greater is the possibility of reducing the cost of operations.

When absolutely no information is available, and yet the survey has to be completed at one single operation, there is no other alternative but to use unrestricted random sampling without zoning, or with such zoning or stratification as may be considered convenient from a purely organizational point of view.

Fortunately, some information is usually available. Also, especially in large-scale work, the survey has often to be continued from year to year or at suitable intervals. It is therefore usually possible to adopt the exploratory method, in which a sample survey is first carried out on a very small scale with the primary object of collecting basic information required for preparing an efficient design for later surveys. Sometimes such preliminary surveys have to be carried out more than once, and the scale of operations is gradually increased until finally the whole area or universe is fully covered. For example, in the Bengal crop survey a pilot survey was first carried out in 1937 covering only 124 square miles at a total cost of about £1,100. In 1938, a second survey was organized, in the light of the experience gained in the previous season, on a larger scale covering about 400 square miles at a cost of £2,500 ; next year the total area covered was nearly 2,600 square miles at a cost of £6,000. In 1940 the area covered was nearly 20,600 square miles at a cost of £8,100; and finally, in 1941 the whole jute tract of Bengal, measuring about 60,000 square miles, was surveyed at a cost of £10,100.

In the exploratory stage the object is to gather as quickly as possible some rough idea of the variance function, cost of operations, type and nature of human agency likely to be suitable for the field work or fluctuations from one region to another (*i.e.*, the distribution over space) of the elements or statistical variates proposed to be estimated in the survey. An intensive study is often essential in the exploratory stage. The cost per square mile (or per unit) is therefore necessarily high, but as basic information begins to be gathered the design of the survey steadily improves, and the cost per unit rapidly decreases in subsequent surveys.

In the Bihar crop survey the exploratory work was done in only two districts covering about 8,000 square miles from February to April 1944. As conditions were found in many ways similar to those in Bengal, it was possible to organize a full-scale provincial crop survey in Bihar covering about 70,000 square miles in the next crop season extending from July to September 1944. This survey was, however, admittedly weak in certain respects. In the next survey in the main winter rice season from October to December 1944 the Bihar crop survey was, however, fully organized. This shows that where previous experience is available it is possible to expedite the exploratory phase.

In extensive surveys it is usually necessary to employ a large number of staff in both field and statistical branches. For example, in the Bengal crop survey the field staff consists of about 350 workers, who actually go round from village to village, with a number of inspectors and supervisors at the top. The total statistical staff in the Institute for computational and technical work consists at present of about 375 workers inclusive of all grades in 1945-46. Giving necessary training to the staff is an important part of the work, and takes a good deal of time. The exploratory approach is often convenient from this point of view.

In socio-economic or demographic enquiries also it has been found extremely helpful to conduct a small-scale preliminary enquiry, and then make final arrangements in the light of the experience gained thereby. A pilot survey enables forms and schedules, staff organization and other details being thoroughly worked out in advance, and often saves its cost many times over.

In large-scale surveys which are repeated from season to season or from year to year the exploratory development of technique, of course, continues all the time. As pointed out by A. Wald, the underlying logic is the same as in sequential tests. Advantage is taken of all available information in preparing the design of the survey. As information accumulates, the efficiency of the survey continues to improve from one survey to another.

Multiple surveys. A special form of the exploratory approach also deserves mention. Consider an actual example. In 1941 it was desired to develop a method for forecasting the yield of cinchona bark in blocks of standing plants each covering from 20 to 40 acres. In the first survey a number of physical measurements, such as girth, height, number of shoots, surface area of the plant, etc., were made on a number of standing plants picked up at random. These plants were then uprooted and the yield of bark was determined for each plant separately. Coefficients of correlation were then determined between the yield of bark and the different physical measurements. Three measurements—namely, height of the plant, girth at a height of 6 inches above ground, and the number of shoots—were selected for final use. A number of plants in the particular block under survey can be then picked up at random and these three physical characters measured for each plant, from which the average values of these three characters for the block as a whole can be determined. The plants are not, however, uprooted, and no direct measurements of bark yield are made during the second survey. (The number of plants in the block can be determined, if necessary, by a subsidiary survey on a grid basis.) Using regression equations based on the coefficients of correlation determined in the first survey, it is then possible to estimate the total yield of bark for the block as a whole.

Such methods have been used, for example, in forestry work in the U.S.A., and have been sometimes called "double sampling." Various other possibilities are of course open; recourse may be had to triple or quadruple or n-ple samples, so that this type of work may be suitably called the method of multiple sampling (which is, of course, to be carefully distinguished from multi-stage sampling). In one sense it is really an extension of the method of concomitant variations. The statistical relation between two (or more) variates is first determined by a preliminary sample survey or a series of such surveys. Estimates are then prepared for one of the variates by a second or subsequent survey, and estimates for the other correlated variates are then calculated with the help of the previously determined regression relations. A simple example occurs in crop-cutting experiments. In the Indian Statistical Institute the weight of green plants of jute or the weight of paddy immediately after harvesting are recorded on an extensive scale. In only a small fraction of cases (of the order of 10 per cent), the jute plant is steeped in water, retted and the dry fibre extracted and its weight determined directly, or the paddy is dried, husked and the weight of rice measured separately. These auxiliary measurements serve to supply the regression relation between the weight of green plants of jute or the weight of paddy immediately after harvesting and the yield of dry fibre of jute or of husked rice, respectively, which can then be used to estimate the corresponding final yields from the more extensive weights taken immediately after harvesting. Such a procedure simplifies the field work enormously without any appreciable loss of precision in the final results.

Such methods, in which the estimates made in later surveys are based on correlations determined in earlier surveys, may perhaps be called "covariate sampling."

PLANNING OF SAMPLE SURVEYS

Coming back to the planning of sample surveys, the ultimate aim from the statistical point of view is, of course, to reduce the over-all or effective margin of error to the desired extent. It has been explained that in large-scale work this involves considerations of time, cost, and the human agency. These three factors are intimately connected, and are merely different facets of the same thing—namely, the plan or design of the survey. Any change in human agency, in general, would involve changes in the time programme, the total cost, and also (a point to be remembered) in the effective margin of error. In fact we learnt quite early the overwhelming importance of the human factor in the organization of statistical sampling on a large scale.

The over-all or effective margin of error does not consist simply of sampling variations, but also of errors of ascertainment arising from the fallibility of the human agency. W. Edwards Deming has emphasized this point on many occasions, and has given a comprehensive review in his paper on "Errors in Surveys" (*American Sociological Review*, Vol. IX, No. 4, August 1944, pp. 359–369). He gives a list of thirteen factors affecting the ultimate usefulness of a survey. I shall adopt a simpler classification for present purposes and omit differences arising from variability in response or changes in the universe, or the population or the field under survey. These sources of error are of particular importance in socio-economic or public opinion surveys and offer fruitful subjects for research. A good deal of work has been already done and is still going on in this field in the U.S.A. and U.K. In the Indian Statistical Institute we have also made a few studies, but on a comparatively small scale, as the great bulk of the Institute work has been concerned primarily with crop surveys, in which these problems are not of great importance.

CLASSIFICATION OF ERRORS

Very early it became clear that mistakes committed in the compilation of primary data and, to a less extent, at different stages of processing the material, were equally or even more important than errors of sampling. Continuous attention has had to be given, therefore, to this aspect of the problem. Faults in planning (including defects in the design and technique of sampling, the preparation of questionnaires and field schedules), as aslo mistakes in the final analysis or interpretation, belong to the purely scientific and technical level of the work, and have been kept out of the present discussion. The remaining sources of inaccuracy have been usually considered in the Institute under the three broad heads described below.

Sampling variations, which have to be considered by familiar methods of theoretical statistics. The large sample theory (based on the normal distribution) has been naturally used most extensively. Observed frequency distributions are, however, sometimes definitely non-normal in character, and special forms have been or are being investigated in certain cases. The logarithmic (for example, in crop yields) or square-root transformations (in certain economic and demographic studies) or truncated distributions (for example, in studying the correlation between weight of crop immediately after harvesting and weight of grain after drying) have been found useful. The subject, however, deserves and requires extensive investigations.

Recording mistakes have been usually considered under two sub-heads: mistakes which occur either through bias or personal equation of the observers or gross negligence or even deliberate dishonesty on their part; and mistakes which arise at the stage of compilation, classification, tabulation, computation, and final analysis and presentation of the material. Both types have been investigated in considerable detail in the Statistical Institute, and a great deal of attention is still being given to developing methods for controlling such recording mistakes.

Physical fluctuations. It has also been found that for every element (or quantity) sought to be estimated by the sampling method there exists a normal margin of physical uncertainty which is a characteristic feature of the element under survey. For example, it is believed that wheat increases in weight by as much as 4 or 5 per cent by the absorption of moisture during the rainy season in a dry province like Sind. Traders, in fact, move the grain by railways at such times that the gain accruing from the increase of weight practically pays for the cost of storage and freight. In this case the exact meaning of the term "total production of wheat" has to be carefully specified. But even when this has been done, a residual margin of physical fluctuations of the order of probably 1 or 2 per cent would still remain. It is clear that no useful purpose would be served by trying to reduce the sampling error to a value below the margin of such physical fluctuations.

CONTROLS AT THE STAGE OF STATISTICAL PROCESSING

Output and mistakes occurring at different stages of copying, compilation, or numerical computations have been studied in considerable detail. For this purpose, computational work has been broken down into a large number of jobs or items with standard specifications such as copying three-figure tables, adding four-figure quantities, squaring three-figure entries, preparing frequency tables with not more than ten classes, etc. For each item, standard rates of output have been worked out, as also permissible limits of rates of mistakes. For each item standard rates of payment have also been set up. A daily job account is kept of the item, time spent, and output for each computer separately, and these are punched on Hollerith cards and tabulated for the valuation of the total work done by each computer in each month.

Individual differences are also studied, which enables inefficient workers to be weeded out on an objective basis, and work to be distributed among different workers in a suitable way. (Some workers were found to have such high rates of mistakes that their net valuation was negative, showing that a premium would have to be paid to the Institute for employing them.) The advantage of the above system is two-fold. It supplies a scientific method for controlling mistakes at different stages of the processing of the primary material. It also supplies a sound basis for cost accounting, which is a characteristic feature of work in the Institute.

In the valuation of computational work it is necessary, of course, to take into consideration not merely the total output, but also the number of mistakes committed. The method of making deduction on account of mistakes is simple in principle. Each mistake involves a certain amount of expenditure in correcting it; and, in theory, the deduction made should be equated to this additional expenditure. In practice the matter is, however, complicated, but the subject is being studied, and attempts are being made to set up standard procedures.

Sub-samples in statistical analysis. In large-scale computational work it has been sometimes found convenient and useful to carry out the statistical analysis in the form of a number of sub-samples. The sample units are allotted consecutive serial numbers in the order in which these are located on maps or picked up for being included in the survey. It is possible therefore to divide the sample data into a number of independent random sets by choosing sample units with odd or even serial numbers, or with numbers ending in particular digits. The actual computational work and analysis are then done separately for each such randomized sub-set of the material. This has three distinct advantages. Firstly, it often enables dimensional results to be obtained very quickly, which is sometimes of great help to administrators. Secondly, a comparison of the results of different sub-sets often gives clues for the detection of gross mistakes in procedure or in calculations. Finally, results for different sub-sets supply a very good idea of the over-all or effective margin of error.

Controls in mechanical tabulation. It may be mentioned here that during the last two years a great deal of the Institute work has been mechanized. The Institute has at present a full equipment of Hollerith eighty-column sorter, tabulator, and multiplier units with a battery of about fifty hand-operated punching and verifying machines, and a summary reproducing punch. In the Hollerith system each card is punched, and then passed a second time through a verifying machine operated by a different worker. With about fifty punching and verifying workers, the human factor becomes quite important. In the beginning there was some difficulty in getting the verifying work done with proper care. A simple but effective statistical control has been now introduced which is probably worth mentioning. A concrete example will make the method clear. Consider the procedure for checking the accuracy of punching of say 10,000 cards. A suitable number (say 100 or 200 or 500) of punched cards are withdrawn from the set at random. A dummy set of an equal number of test cards are punched deliberately with mistakes in assigned columns; and this second set is substituted for the first set at the appropriate places. The whole set of 10,000 cards is then verified. It is clear that, if the work is done properly, then the set of (100, 200 or 500 as the case may be) test cards must be rejected at the stage of verification in the exact order of their serial number. Besides the test cards it is possible and likely that other cards also would be rejected; but the test cards must come out in any case. This supplies a good foolproof check at the stage of verification.

Efficiency at the stage of processing. Generally speaking, the basic problem may be stated in the following way. A given volume of computational work (in the most general sense, including scrutiny of primary material, sorting, tabulation, statistical analysis, checking, etc.) is required to be done. The problem is to organize the processing work in such a way that it can be completed with maximum accuracy in minimum time and at minimum cost. The subject is important and fascinating, and attempts are being made to study it as far as possible along with the regular work of the Indian Statistical Institute. Valuable information has been collected and indications for future investigations have been secured, but it has not yet been possible to tackle the problem on fundamental lines for lack of resources.

ILLUSTRATIVE EXAMPLES

I shall now try to give some idea of the kind of information which it has been possible to collect by the method of statistical sampling in recent years. It is quite impossible, of course, to give a detailed account of the numerous enquiries undertaken by the Indian Statistical Institute during the last five or six years. I have therefore made a selection of typical examples which are likely to be of interest outside India. The use of replicated networks of samples (in which information for each network is collected independently by different sets of investigators) has been a characteristic feature of the work in the Institute. It includes as a special case the same sample being surveyed twice or more often by independent investigators. Such methods have been used very successfully in furnishing estimates of the over-all margin of error, and I have intentionally laid stress on the use of replication in choosing my examples.

CROP SURVEYS

This is the field in which the largest volume of work has been done in the Indian Statistical Institute. An account has been given elsewhere (*Philosophical Transactions*, B 584) of the general development of work which took place from 1937 to 1941, in which year a sample survey of the area under jute was carried out throughout the province of Bengal. The survey of jute acreage was repeated in 1942. Although by that time, owing to Japan's entry into the war, the food situation in Bengal had already become difficult, I failed completely to persuade the Government to extend the sample survey to cover the paddy crop in Bengal. The Bengal famine occurred in 1943. Since that year we have had the opportunity of carrying out a sample survey of both jute and rice crops throughout the province.

Bihar Crop Survey. As already mentioned, a pilot survey had been started in only two districts, comprising about 8,000 square miles, in the *rabi* (winter wheat) season in February 1944. The scheme was rapidly expanded to cover the whole province, and a full-scale survey of *aghani* (winter rice) of 1944-45 season was carried out only a few months later.

TABLE 1. BIHAR CROP SURVEY: PROGRESSIVE REPORTS ON WINTER RICE, 1944-45 ESTIMATED AREA (IN HUNDRED THOUSAND ACRES) UNDER DIFFERENT CROPS

serial no.	field work upto		grids surveyed		sampling fraction	area in hundred thousand acres under			
			number	percentage	one in	rice	pulses	sugarcane	potatoes
(1)	(2)		(3)	(4)	(5)	(6)	(7)	(8)	(9)
1	October	17th	2,588	4.0	4175	133.35	29.55	—	—
2	November	2nd	14,361	22.3	752	135.37	30.65	3.72	—
3	,,	11th	19,813	30.8	545	140.70	31.28	3.74	—
4	,,	18th	25,934	40.4	417	141.71	30.31	3.94	—
5	,,	26th	32,312	50.3	334	140.89	29.77	3.87	—
6	December	4th	38,818	60.4	278	139.46	29.78	4.01	—
7	,,	12th	45,006	70.0	240	138.72	30.40	4.03	—
8	,,	21st	51,502	80.1	210	140.05	29.63	4.06	—
9	January	12th, 1945	64,260	100.0	168	136.50	29.03	3.97	1.12
			(total area covered = 432.23) proportion			= 31.58%	6.72%	0.92%	0.26%

The above table shows the actual progressive estimates which were submitted to the Government from time to time. It will be noticed that the first estimates were based on only 2,588 grids (or about 4 per cent of the total number ultimately surveyed), which represented a sampling fraction of one in 4,175, and yet the estimates of the area sown with rice (133.35) and with pulses (29.55) were of right dimensional order and compare quite favourably with the final figures, 136.50 and 29.03 (in 100,000 acres), respectively. After receiving the first estimates of rice and pulses, the Government desired to have similar figures for the area under sugarcane, which were submitted regularly from the second progressive report. At the last stage, at the request of the Government, an estimate of 1.12 (in 100,000 acres) was supplied for the area under potatoes.

The total geographical area covered in the survey was 432.23 (in 100,000 acres), so that the proportion of land was 31.58 per cent under rice, 6.72 per cent under pulses, 0.92 per cent under sugarcane, and only about 0.26 per cent, or one quarter or 1 per cent, under potatoes. This would supply a good idea of the wide range of the survey. It, however, scarcely needs mentioning that the percentage (but not the absolute) error would be naturally quite high for potatoes or sugarcane which have comparatively small acreages.

Another point is worth explaining. Sowings of rice continue for a good length of time, possibly till the middle of November, depending on actual conditions of rainfall in different areas. In Bihar (as well as in Bengal) the rice crop is mainly rain-fed, and usually a certain proportion of marginal land is put under cultivation in the hope of a favourable distribution of rainfall. When the rainfall is not exceptionally favourable (as often happens) it is inevitable that some of the lands which in the earlier part or middle of the season are under a particular crop will go out of cultivation owing to lack of rain at the end of the season. The total area under rice therefore increases from the beginning to the middle of the season, and then usually decreases. For this reason the earlier estimates would often agree fairly well with estimates based on material collected at the end of the season.

Bengal Crop Survey. A brief description of the Bengal Crop Survey would be useful at this stage. The province of Bengal has been divided into about 1,100 "cells", each approximately 64 square miles in area. In doing this the boundaries of *mauzas* (broadly corresponding to villages, and roughly about one square mile in area) have been kept intact, so that each *mauza* (or village) lies entirely within a single cell. This has been done in order to facilitate the building up of estimates for higher administrative or revenue units such as *thana* (or police circle of 150 square miles on an average), sub-division (about 850 square miles) or districts (about 2,500 square miles), and finally the province as a whole (70,000 square miles).

Flexibility of the design. The design of the sample survey has naturally to take into consideration the intensity of cultivation or the importance of particular economic variates desired to be investigated in different regions. For example, in crop surveys it is legitimate to omit areas known to be under forests. In the case of particular crops like jute it is permissible and necessary to allot a comparatively small number of sample units to regions or cells where jute is known to be grown only in small quantities (when it is desired to obtain an accurate estimate of the total production of the crop for the province as a whole). The arrangement by which the whole province is divided into a large number of cells gives great flexibility in designing, as it permits unimportant cells to be left out of the survey altogether or to be allotted a comparatively small number of grids. The first stage in the design of the survey is thus to settle the size (area) and the number of grids to be allotted to the different regions or cells. For surveys covering a number of crops or a number of different elements it is often convenient to adopt a uniform density, or the same number of grids for each cell included within the survey. But this is a special case; in general, the number in each cell can be varied as desired.

In Bengal the total number of sample units allotted to each cell (of 64 square miles) is at present of the order of 100 for crop surveys. The position of each grid is located at random and marked on maps. In this way each *mauza* gets none or one or more random points. These random points are numbered serially as they are marked on the map, and a grid (of 2.25 acres for crop surveys) is located at each random point. All fields falling within the grid are subsequently surveyed by the field staff.

Two sub-samples. Within each cell (in which the act of randomization is carried out separately) the sample units, as they are located at random, are given consecutive serial numbers, so that the set of sample units or grids with even serial numbers and the set of sample units or grids with odd serial numbers form two independent but interpenetrating and random networks of sample units, each of which covers the same area. These two sets are allotted to different parties of field investigators, so that information for each set is collected independently. These two sets of data thus supply two independent estimates for each cell as a whole.

Duplicate grids. A further control is used in crop surveys as a routine measure at present. A certain number of grids or sample units are allotted to both networks so that information for such "duplicated" sample units is collected by both sets of field investigators. In the Bengal Crop Survey in 1946, for example, out of fifty-four grids allotted to each network of samples, fourteen are "duplicated" or common to both sets. A detailed comparison of the two sets of field records for these fourteen grids in each cell shows how far the field survey has been done with care. The field investigators know that certain grids are duplicated, but have no knowledge as to which particular grids have to be enumerated twice. The field programme is arranged in such a way that the two different parties of investigators are separated by at least two or three days' journey, so that there is no chance of the investigators copying the records from one another. This method has proved extremely useful in supplying a good control at the point of collection of the primary material.

It is worth mentioning briefly another control used in connection with duplicated (or sometimes replicated) grids or sample units which are enumerated twice (or more often) by different parties of investigators. In the crop survey the investigators are required to note the average height of the plant in each field. As the duplicated (or replicated) surveys are done at different times—sometimes with an interval of even two or three weeks—this supplies a good check. For example, cases have occurred in which the height of plants were recorded as being lower at the time of the later survey than the height recorded during an earlier survey. Such discrepancies can arise only through mistakes in identifying the fields, or through gross negligence or dishonesty on the part of the investigators. Such controls also serve the useful purpose of supplying an objective basis for the rejection of unreliable primary material.

Agreement between half-samples. Table 2 shows the two independent (replicated) estimates of crop acreages in Bengal during the three years 1943, 1944 and 1945. The design was that of two interpenetrating random networks of samples. The actual design of the survey was different in each year, but it is not possible to enter into these details. The two independent samples (*A* and *B*) are usually called half-samples as both together make up the complete sample. These might also have been called "double" samples; but this would have probably raised misapprehensions about cost in the minds of administrators.

TABLE 2. BENGAL CROP SURVEY : HALF-SAMPLE COMPARISONS, 1943–45

year	number of grids surveyed			area under crops in thousand acres			difference ($A-B$)		standard error of difference	Fisher's t
	sample (A)	sample (B)	combined	sample (A)	sample (B)	combined	actual	percentage		
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
				jute crop						
1943	29,676	29,676	59,352	2,759	2,757	2,758	2	0.1	100	0.02
1944	30,487	30,037	60,524	2,150	2,056	2,106	94	4.45	50	1.87
1945	53,504	52,623	106,127	2,512	2,528	2,520	— 16	0.8	*	—
				aus (monsoon) rice						
1943	29,676	29,676	59,352	6,807	6,923	6,865	—116	—1.7	446	0.26
1944	30,457	30,037	60,524	7,815	7,942	7,873	—127	—1.6	90	1.41
1945	53,504	52,623	106,127	6,966	6,805	6,864	161	2.3	*	—
				aman (winter) rice						
1943	31,216	31,215	62,431	23,840	24,044	23,942	—204	—0.8	703	0.29
1944	50,501	50,107	100,608	22,491	21,903	22,201	588	2.7	550	1.07
1945	46,148	45,971	92,119	20,970	21,202	21,087	—232	—1.1	103	2.26†

* Error calculations for 1945 were not completed at the time of writing.

† Significant at 5 per cent level.

The two crop estimates in thousand acres based on the two independent samples (A and B) are shown in cols. 5 and 6 respectively, and the pooled estimate in col. 7. The difference between the two estimates $A - B$ in thousand acres is shown in col. 8 and the same difference expressed as percentage of the combined value given in col. 7 is shown in col. 9. The standard error of the difference calculated from the samples is given in col. 10, and Fisher's t in col. 11. This table shows the kind of results which are actually obtained in practice. I should like to note, however, that the field survey in 1943 had to be organized at very short notice and under great difficulties (owing to the incidence of the Bengal famine), and supervision was probably not quite adequate; the close agreement between the two half-samples may be, therefore, to some extent spurious.

Discrepancies in crop enumeration. Since 1937, when work was started by the sampling method on a very small scale, the relative accuracy of the results obtained by a so-called complete enumeration and by the sample survey respectively have been the subject of acute controversy in India. From the very beginning attempts were therefore made to investigate the reliability or otherwise of crop enumeration done by ordinary field investigators in actual practice. The method adopted has been to carry out independently twice (or more often) complete enumerations of crops on fields in the same village or region by entirely different sets of investigators. If crop enumeration can be carried out with complete accuracy, then the two sets of records should be in perfect agreement. Certain discrepancies are, however, introduced by the fact that the two surveys have necessarily to be carried out one after the other in order that there is no chance of one party of investigators copying from the records prepared by the other party. If the first survey is carried out before sowings are completed, it is possible that certain fields which were uncultivated at the time of the first survey might have been brought under cultivation by the time the second survey is made. Or if the second survey is made rather late in the season, it is possible that some of the fields having standing crops at the time of the first survey might go out of cultivation at the time of the second survey for lack of rainfall or other causes. Such discrepancies can, however, be usually distinguished from a knowledge of crop conditions in relation to the two dates of survey.

A great deal of material of this nature has been compiled, year by year, since 1937. Even a cursory examination of the material leaves no doubt that very large discrepancies occur at the stage of crop enumeration. Some of the mistakes are genuine, in the sense that the investigators actually fail to distinguish between different crops, or make wrong estimates of the proportion of the field under different crops. Mistakes due to gross negligence, however, also occur; and sometimes, of course, entries are made by pure guesswork by unreliable investigators who want to avoid the trouble of going round to the fields. It must be remembered that conditions in which the crop enumeration has to be done in Bengal are particularly difficult. In summer the temperature would often go up to 110°F. or more; during the monsoon many of the roads are submerged; throughout the crop survey a large number of investigators suffer from malaria and other diseases every year. This makes it all the more necessary to have adequate controls at the point of collection of the material.

I am giving below a few tables to indicate the kind of material collected and nature of comparisons made in the Indian Statistical Institute on an extensive scale from year to year.

TABLE 3. ILLUSTRATIVE SUMMARY COMPARISONS OF DUPLICATE CROP RECORDS PREPARED BY DIFFERENT INVESTIGATORS (A AND B)

province and zone under survey	years of survey	name of crop	estimated acreage		discrepancies ($A-B$) between the two records				percentage to A	
					actual		sum			
			(A)	(B)	positive	negative	algebraic	absolute	algebraic	absolute
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Bengal 2 *thanas*	1937	jute	355	385	103	133	—30	236	—8.5	66.5
Bengal 8 *thanas*	1938	jute	3,584	3,646	466	528	—62	994	—1.7	27.7
Bengal 6 *thanas*	1944	wheat	286	298	100	112	—12	212	—4.3	74.6
Bihar 6 *thanas*	1944	gram	730	636	244	150	94	394	12.9	53.9
Bihar 86 *thanas*	1944	*bhadoi* paddy	6,715	6,481	1,658	1,424	234	3,082	3.5	45.9

The above gives a comparison of duplicated crop records prepared independently by two different sets of investigators called A and B. Cols. 4 and 5 give the two independent figures of total acreage under the same crop in the same geographical area. A detailed comparison of the two sets of records was carried out for each field separately. For example, a particular field may be shown as being under jute by the first party of investigators (A), but not under jute by the second party of investigators (B). In such cases the discrepancies $A-B$ may be considered to be positive. When a field is shown under a particular crop by party B, but not by party A, the discrepancy $A-B$ would be then naturally considered to be negative. If such discrepancies are added up for each individual plot, one would get the total positive or negative discrepancies. These are shown in cols. 6 and 7, and the point to be emphasized is that such absolute discrepancies are actually very large, and in the above table range from 27.7 per cent to 74.6 per cent. So far as detailed crop records are concerned, it is quite clear, therefore, that the so-called complete enumeration cannot supply reliable results under conditions actually existing at present in Bengal or Bihar.

Many of the mistakes are, however, unbiased and tend to cancel out. The total algebraic errors are obtained by simply adding the positive and negative discrepancies, and are shown in col. 8. For example, in the first line of Table 3 the figure—30 in col. 8 is obtained by adding the positive and negative discrepancies in cols. 6 and 7, and is of course also simply the difference between the two figures given in cols. 4 and 5 respectively. In comparison with the absolute discrepancies shown in col. 9 it will be seen that algebraic discrepancies are proportionately much smaller, owing to the cancelling out of positive and negative mistakes. In other words, results of complete enumeration of crops as carried out at present in Bengal or Bihar have exactly the same qualitative kind of statistical reliability as results based on sample surveys, with the important difference that in the so-called complete enumeration (as usually conducted without any superposed sampling controls) there is no possibility of estimating the margin of error. This is a conclusion of the greatest practical significance.

Two illustrative examples of a detailed comparison, plot by plot, of duplicated crop records are given below. Table 4 refers to material collected in *mauza* (village) Kazikandi in P.S. Goalundo (an administrative unit in Bengal in 1943). In this village 332 fields were surveyed by investigator A on August 14th, 1943. He noted the name of the crop (or mixture of crops) growing on each individual field. A second complete enumeration was made by investigator *B* about a fortnight later, on September 2nd, 1943, who also noted independently the name of the crop or mixture of crops growing on each individual field. Two records were thus available for each field, and can be shown in the form of a two-way table as in Table 4.

TABLE 4. BENGAL CROP SURVEY, 1943: COMPARISON OF DUPLICATED COMPLETE ENUMERATION IN P.S. GOALUNDO, *MAUZA* KAZIKANDI

		B-survey September 2nd, 1943				
		jute	*aman* rice	jute–*aman*	no crop	total
A-survey August 14th, 1943	jute	4	15	4	3	26
	aus rice	4	12	1	4	21
	jute–*aus*	17	66	2	9	94
	jute–*aus*–*aman*	—	2	—	—	2
	aus–*aman*	1	—	—	—	1
	no crop	37	45	4	102	188
	total	63	140	11	118	332

If the crop enumerations made by the two investigators, (*A* and *B*) were in complete agreement, then all fields shown as being under jute by investigator *A* would also be shown under jute by investigator *B*, and vice versa, so that entries would occur only in the diagonal cells. A glance at Table 4 would show that the actual position was entirely different. For example, according to investigator *B* there were altogether 63 fields under jute on September 2nd. According to investigator *A*, who had surveyed the same identical field a fortnight earlier, apparently only four were growing jute, four *aus* (monsoon) rice, seventeen had a mixture of jute and *aus* rice, one field had mixture of *aus* (monsoon) and *aman* (winter) rice while no fewer than thirty-seven fields were without any crops at all. Discrepancies are equally striking for other crops. In fact out of 332 fields surveyed, the two investigators were in agreement only in regard to four plots, which both recorded to have been under jute, and also in regard to 102 fields, which were shown by both as having no crops at all. In other words, in only 106 out of 332 entries was there agreement between the two independent crop enumerations. This is clearly unsatisfactory, and shows that either one or both the investigators had failed to do their work with reasonable care.

TABLE 5. BENGAL CROP SURVEY, 1943: COMPARISON OF DUPLICATED COMPLETE ENUMERATION IN P.S. CHUADANGA, *MAUZA* MANIRAMPUR

		B-survey August 10th, 1943			
		aus	jute-*aus*	none	total
A-survey June 28th, 1943	jute	1	—	1	2
	aus	123	2	8	133
	jute-*aus*	—	3	—	3
	none	7	—	189	196
	total	131	5	198	334

TABLE 6. BIHAR CROP SURVEY, *BHADOI*, 1944-45: COMPARISON OF TWO INDEPENDENT ESTIMATES OF P (PROPORTION OF LAND UNDER *BHADOI* RICE IN EACH INDIVIDUAL PLOT) BY TWO PARTIES OF INVESTIGATORS

(District, Hazaribagh; Thana, Giridih; No. of villages, 86)

anna estimates of proportion of land under rice during 2nd survey between Aug. 6th and Oct. 12th	first survey / second survey	anna estimates of proportion of land under rice during first survey between August 3rd and September 29th, 1944																	
		zero	1-anna	2-anna	3-anna	4-anna	5-anna	6-anna	7-anna	8-anna	9-anna	10-anna	11-anna	12-anna	13-anna	14-anna	15-anna	16-anna	total
	zero	23,988	72	116	29	116	26	51	14	108	10	33	13	68	15	31	24	4,307	29,021
	1-anna	87	6	8	3	2	—	1	—	3	—	—	1	1	—	—	3	33	148
	2-anna	115	6	12	3	4	—	1	2	9	2	—	—	2	1	—	2	79	238
	3-anna	32	6	3	—	2	—	—	—	3	—	5	—	1	—	—	—	24	76
	4-anna	129	6	9	2	16	1	7	1	9	—	4	1	2	—	2	—	102	291
	5-anna	14	1	1	1	—	—	—	—	1	—	1	—	1	—	—	—	17	37
	6-anna	59	—	3	1	6	—	3	1	5	1	—	3	2	2	—	—	56	142
	7-anna	23	—	—	—	3	—	—	—	1	—	1	—	—	—	—	—	19	47
	8-anna	130	4	3	1	7	—	5	—	13	1	7	1	4	—	5	1	157	339
	9-anna	9	—	1	—	2	—	2	—	1	—	1	—	—	—	—	—	13	29
	10-anna	44	—	3	3	3	—	2	—	3	1	8	2	3	1	2	—	77	152
	11-anna	18	—	—	—	2	2	1	—	—	—	—	—	—	—	—	—	13	36
	12-anna	63	—	—	1	5	—	3	1	6	—	1	—	4	1	2	—	109	196
	13-anna	7	1	—	—	—	—	—	1	1	—	1	—	—	—	1	—	26	38
	14-anna	43	—	2	—	—	—	—	—	5	—	2	—	4	—	1	3	112	172
	15-anna	21	—	—	—	3	—	1	—	2	—	1	—	3	3	4	1	121	160
	16-anna	3,721	27	25	11	65	10	43	10	113	18	70	16	82	34	109	82	20,635	25,071
	total	28,503	129	186	55	236	39	120	30	283	33	135	37	177	57	157	116	25,900	56,193

A more satisfactory example is shown in Table 5 for another pair of investigators working in Bengal in the same year. In this case out of 131 fields recorded as having *aus* (monsoon) rice by investigator B on August 10th, 1943, no fewer than 123 had been shown as having rice also by investigator A during the earlier survey on June 28th, 1943. Out of five fields shown as having both jute and *aus* (monsoon) rice by investigator B, three were shown so in the earlier survey, while two fields were shown as having only *aus* rice by investigator A. This is quite possible, as *aus* rice is sown earlier, and these two fields might have been sown with jute after the first survey was over. As regards uncultivated fields, the two investigators had agreed in the case of 189 out of 196 fields. There was agreement in 315 (or 94.3 per cent) out of 334 fields. This may be contrasted with the agreement in the case of only 106 (31.9 per cent) out of 332 fields in the previous table. Such differences show the wide variation in the quality of field work done by different investigators. The importance of providing statistical controls like duplicated grids can be therefore easily appreciated.

I may now give a quantitative example. In the sample survey of crops the field staff is required to estimate the proportion of land in each field which is under a particular crop. These estimates are made in terms of the Indian coin anna, sixteen of which make up a rupee. In other words, crop estimates are made in units of $6\frac{1}{4}$ per cent.

Table 6 shows a comparison of results of such estimates of the proportion of land sown with rice on the same group of 56,193 fields which were surveyed independently by two sets of investigators. The comparison between the two sets of records has been made with increasing latitude of discrepancy. For example, if the two entries (estimates of the proportion of field under rice) relating to the same field (made by the two different parties of investigators) agree within 1 anna, (or $6\frac{1}{4}$ percent), then the latitude of comparison is 1 anna, or $6\frac{1}{4}$ per cent. In the same way, if the two entries agree within 4 annas, then the latitude of comparison is 25 per cent; and if they agree within 8 annas, the latitude is 50 per cent. Finally for a 16-anna (or 100 percent) latitude of comparison two entries relating to the same field would be considered to be in agreement if both the parties record the field to be under rice irrespective of the quantitative proportion.

TABLE 7. BIHAR CROP SURVEY, 1944: *BHADOI* (MONSOON) RICE: COMPARISON OF COMPLETE ENUMERATION BY TWO SETS OF INVESTIGATORS

latitude of comparison	number of plots surveyed: in agreement: no crop	number of plots surveyed: in agreement: with rice	number of plots surveyed: not in agreement	total number of plots compared	percentage of plots surveyed: in agreement: no crop	percentage of plots surveyed: in agreement: with rice	percentage of plots surveyed: not in agreement
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1 anna ($6\frac{1}{4}$%)	23,988	20,699	11,506	56,193	42.69	36.84	20.47
4 anna (25%)	23,988	21,269	10,936	,,	,,	37.85	19.46
8 anna (50%)	23,988	21,742	10,463	,,	,,	38.69	18.62
16 anna (100%)	23,988	22,657	9,548	,,	,,	40.32	16.99

A glance at the summary Table 7 would show that if a $6\frac{1}{4}$ per cent latitude of comparison is permitted, then records relating to 11,506 (or 20.47 per cent) of 56,193 fields were discrepant. Increasing the latitude of comparison to 25 per cent, the number of discrepant fields is only slightly reduced to 10,939, or 19.46 per cent. Allowing a much higher margin of comparison of 50 per cent, the number of discrepant fields is again only very slightly reduced to 10,463, or 18.62 percent. Finally, allowing the maximum possible latitude of comparison—that is, considering merely whether the field is recorded to have rice or not—it is seen that no fewer than 9,548 or 16.99 per cent of fields still show discrepant entries. The latitude of comparison thus makes very little differences, which shows that mistakes usually occur in the identification of the fields rather than in making quantitative estimates of " p " (the proportion of land under a crop).

Comparison of "duplicate" grids. It has been already mentioned that in the same survey of crops the present practice is to have a certain proportion of grids enumerated in duplicate by two independent sets of investigators. This supplies a valuable check on the quality of the field survey. The field investigators make an entry for each grid of the proportion (called "p") of the total land included within each grid (namely, 2.25 acres in the Bengal Crop Survey of 1945-46) which is estimated to be under *aman* (winter) rice. Such estimates were prepared twice by independent parties of field staff so that for each grid there are two values of "p", one estimated by party A and the other by party B. The two sets of records can therefore be shown in the form of a two-way table as in Table 8.

TABLE 8. BENGAL CROP SURVEY, 1945-46 : COMPARISON OF TWO INDEPENDENT ESTIMATES OF P (PROPORTION OF GRID UNDER *AMAN* RICE) BY TWO PARTIES OF INVESTIGATORS

		half-sample (B)												
	p (in%)	0	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80	81–90	91–99	100	total
half-sample (A)	0	1,159	84	50	33	30	17	18	15	27	24	28	62	1,547
	1–10	74	150	39	22	13	9	11	11	4	14	3	13	363
	11–20	31	35	88	14	14	12	10	11	9	3	5	16	248
	21–30	26	20	22	65	18	18	16	6	5	7	6	8	217
	31–40	15	10	16	25	97	42	15	15	13	16	11	11	286
	41–50	17	11	7	14	31	85	32	14	17	12	14	9	263
	51–60	17	6	9	8	18	29	71	30	24	11	16	30	269
	61–70	22	6	7	6	13	10	27	78	37	31	22	25	284
	71–80	24	7	7	8	11	23	17	45	91	45	32	45	355
	81–90	28	7	9	11	9	9	25	15	45	129	52	57	396
	91–99	29	4	7	6	13	11	12	17	29	48	264	143	583
	100	68	13	11	14	13	23	23	31	54	64	152	927	1,393
	total	1,510	353	272	226	280	288	277	288	355	404	605	1,346	6,204

Frequency constants : mean $p(A) = 52.0$ percent, mean $p(B) = 51.9$ per cent. s.d. of $A = 40.9$ per cent, s.d. of $B = 41.2$ per cent, coefficient of correlation $= 0.739$.

TABLE 8a. SUMMARY TABLE OF PROPORTION OF AGREEMENT WITH DIFFERENT LATITUDES OF COMPARISON

	number of grids	accumulated totals	accumulated percentage
complete agreement	3,204	3,204	51.6
1–10% margin	1,069	4,273	68.9
11–20% ,,	511	4,784	77.1
21–30% ,,	367	5,151	83.0
31–40% ,,	225	5,376	86.7
41–50% ,,	185	5,561	89.6
51–60% ,,	142	5,703	91.9
61–70% ,,	96	5,799	93.5
71–80% ,,	106	5,905	95.2
81–90% ,,	86	5,991	96.6
91–99% ,,	83	6,074	97.9
100% ,,	130	6,204	100

If the two sets of crop enumeration made by the two different parties (*A* and *B*) were in complete agreement, then the entries would occur only in the diagonal cells and the coefficient of correlation between the two sets of records would be +1. This of course cannot happen in practice, as different investigators would have different "personal equations" of observation and estimation. In the table shown above it would be noticed that in 3,204 (or 51.6 per cent) out of 6,204 grids, the two sets of records are in complete agreement. If agreement is defined to include a margin of variation upto 10 per cent on either side then 4,273 (or 68.9 per cent) of all grids are in agreement. In the same way about 83 per cent of the grids would be in agreement if a latitude of comparison upto 30 per cent is permitted. This is not unsatisfactory.

The agreement between two sets of records can also be expressed in the form of a coefficient of correlation of +0.739. Owing to the cancellation of positive and negative errors the two mean values $p(A) = 52.0$ per cent and $p(B) = 51.9$ per cent are in entirely satisfactory agreement. Duplicated readings thus show two things—namely, (*a*) detailed agreement, plot by plot or grid by grid, can never be attained in practice even when the field work is done with reasonable care; and yet, (*b*) the agreement between mean values may be quite satisfactory so that reliable sample estimates can be made with confidence.

BIAS IN THE ESTIMATION OF CROP YIELDS

It may be mentioned that in crop-cutting work the principle of random sampling was explicitly recognized for the first time in the classical experiments of J. (now Sir John) Hubback of the I.C.S. in India between 1923 and 1925. He used sampling units of a very small size, 1/3,200 acre, in the form of a triangular cut obtained by a special implement devised by him. He was the first person to point out that owing to high correlation between yields in adjoining parts of the same field there was not much gain in precision by using sample units of a large size. He further emphasized the convenience and economy of using small cuts, of which a much larger number can be collected in the same time by investigators moving from one place to another. He gave a general account of his work in a paper on "Sampling for Rice Yield in Bihar and Orissa" published by the Government of India in

1927 as Bulletin No. 166 of the Imperial Agricultural Institute, Pusa. C. D. (now Sir Chintaman) Deshmukh, of the I.C.S., used Hubback's method a little later between 1928 and 1931 in his crop-cutting work on rice conducted together with settlement operations in two districts in the Central Provinces. About the same time, P. S. Rau, I.C.S., used the same method in two other districts of the same province. I succeeded recently in recovering the original material collected by C. D. Deshmukh, but all papers relating to P. S. Rau's work appear to have been destroyed. As Hubback's paper has been out of print for a long time, it has been reprinted in *Sankhyā* (the Indian Journal of Statistics), Vol. 7, Part 3, pp. 281–94, and in the same number I have given a brief account of Deshmukh's work. In 1938-39 H. P. V. Townend (also of the I.C.S.) used cuts of a comparatively small sampling unit (27.04 sq. ft.) in the form of a square, harvested with the help of a rigid frame, in his work on rice in certain districts in Bengal. It is interesting to observe that four administrators belonging to the I.C.S. used cuts of a small size evidently because they were guided by reasons of convenience and economy.

It is worth mentioning that R. A. Fisher's crop-cutting work on wheat at Rothamsted was influenced by that of Hubback, as stated by Fisher in a memorandum on crop estimating surveys submitted by him to the Imperial Council of Agricultural Research in India on March 2nd, 1945. Subsequently, Yates and others used sample cuts of a small size.

When we first started work on the estimation of crop yields in 1939, we naturally thought of using cuts of a small size. At the same time I thought it advisable to investigate whether the size of the sample unit had any influence on the results. It was necessary at this stage to distinguish between two different possible effects. If the size of the sample unit is appreciable in comparison with the size of the field, then there will be obviously a chance of over-sampling the central area of the field. I have discussed such over-sampling elsewhere (in the paper "On Large-Scale Sample Surveys," *Phil. Trans.* Vol. 231 (B), No. 584, Appendix 6, p. 404). There is a second possibility of bias arising from bordering plants being included within the sample cut. It was this second effect which I had in mind in undertaking an experimental study of the problem.

In our very first series of crop-cutting experiments on jute in Bengal in 1939 we collected some material with sampling units of five sizes ranging between 5.5 sq. ft. and 66 sq. ft. There was some evidence of over-estimation with small cuts, but the material was meagre. In 1940 we used five different sizes of sample cut. A square of size 16 ft. by 16 ft. was located at random in each field (which also was selected at random) and the crop was harvested in the form of a number of separate sub-cuts supplying yield rates for sizes 1 ft. × 1 ft., 3 ft. × 3 ft., 12 ft. × 4 ft., 12 ft. × 12 ft., and 16 ft. × 16 ft. Cuts of different sizes were intentionally obtained from the same spot so as to secure effective local control for purposes of comparison. Information was collected for many items, such as number of plants per acre, the weight of green plants immediately after harvesting, and also the final yield of dry fibre per acre (after the plants were retted, dried, and the fibre extracted). I am giving in Table 9 figures for one particular item—namely, the yield of jute expressed as the weight of green plants (immediately after harvesting) in pounds per acre collected from four different districts in Bengal in 1940.

Yield rates based on 1 ft. × 1 ft. (1 sq. ft.) sample cuts were clearly over-estimates. Such extremely small sizes were therefore omitted in subsequent years. The next larger

TABLE 9. BENGAL CROP SURVEY, 1940-41: YIELD OF JUTE (GREEN PLANT) IN POUNDS PER ACRE

size of cut (sq. ft.) (1)		Dacca ($n = 80$) (2)	Mymensingh ($n = 80$) (3)	Rangpur ($n = 94$) (4)	Tipperah ($n = 66$) (5)	combined ($n = 320$) (6)
1	(1 × 1)	36,603 ± 2,288	24,893 ± 2,535	23,453 ± 1,860	24,292 ± 2,370	27,271 ± 1,160
9	(3 × 3)	20,466 ± 1,317	15,462 ± 872	15,800 ± 749	18,598 ± 1,259	17,462 ± 535
48	(12 × 4)	10,188 ± 576	20,145 ± 774	16,293 ± 650	19,997 ± 848	16,080 ± 411
144	(12 × 12)	9,784 ± 494	19,305 ± 905	18,038 ± 617	20,309 ± 856	16,763 ± 428
256	(16 × 16)	9,332 ± 593	18,836 ± 774	17,552 ± 617	20,433 ± 790	16,828 ± 395

size 3 ft. × 3 ft. (9 sq. ft.) also appeared to give over-estimates, but this size was retained for purposes of comparison. Since 1940 similar material has been collected year by year, for a number of crops like jute, wheat, *aus* (monsoon), and *aman* (winter) rice in different provinces of India. An account of the first series of experiments on jute in 1940 was given in a report submitted to the Indian Central Jute Committee, which had financed the work. The report was printed by this Committee for official use, but was not released to the public owing to war-time restrictions. Other reports on crop-cutting work were submitted by us in subsequent years, but the Government restricted these also for official use. Although a good deal of material has been accumulated, it has not yet been possible to publish a comprehensive account.[1] I am giving a few typical results in Table 10.

TABLE 10. CROP SURVEY: PERCENTAGE YIELD RATES BASED ON SAMPLE CUTS OF DIFFERENT SIZES

size of cut (sq. ft.) (1)	Bengal jute, '40 320 (2)	Bengal jute, '41 185 (3)	U.P. wheat, '41 178 (4)	U.P. wheat, '42 346 (5)	Bengal rice, 1943–44 40 (6)	average index (unweighted) (7)
9	103.8	116.1	121.4	118.1	113.3	114.7
18	—	—	111.6	109.4	—	110.5
25, 27	—	100.7	—	109.0	—	105.0
36	—	—	100.1	99.0	112.1	103.7
48, 49	95.5	95.3	—	—	—	95.4
54	—	—	—	96.0	—	96.0
64	—	105.9	—	—	—	—
81	—	—	—	93.6	—	—
135	—	—	99.9	—	—	—
144	99.6	96.4	—	—	101.2	—
225	—	*100.0*	97.4	—	—	—
256	*100.0*	—	—	—	—	—
324	—	—	—	*100.0*	—	—
576	—	—	*100.0*	—	*100.0*	—

In this table the results have been expressed as percentages of the yield rate based on sample cuts of the largest size used in different series of experiments. On the whole, the above table indicates that the bias decreases as the size of the cut is increased from 9 sq. ft. to probably something of the order of 40 or 50 sq. ft. and becomes negligible with cuts of larger size.

The results given in Table 10 all refer to sample cuts which were located in the field with pegs and ropes. Experiments were also made with rigid and semi-rigid frames made of wood or of wood and iron. I have no time here to discuss these results, beyond stating that there was evidence of over-estimation with small cuts in such cases also.

---

[1] A few observations have been made in *Sankhyā*, Vol. 7, part 3, April 1946, pp 269-80.

I may mention at this stage that in the paper "On Large-Scale Sample Surveys" written at the end of 1942, I referred to the over-estimation arising from the use of sample cuts of small sizes, and explicitly stated (*Phil. Trans.*, Vol. 231 (B), No. 584, p. 409):

> "It was found that there was persistent over-estimation in working with units of very small size. In the case of field survey the obvious explanation is that the investigator has a tendency to include rather than to exclude plants or land which stand near the boundary line or perimeter of the grid. This boundary effect naturally becomes less and less important as the size of the grid is increased. In crop-cutting work on jute it was found, for example, that mean values for all the characters studied (such as number of plants per acre, weight of green plants, weight of dry fibre) were much higher for sample units of small size, so that it was not at all safe to work with cuts of a size less than say 25 sq. ft."

TABLE 11. BENGAL CROP SURVEY, 1945–46 : YIELD IN POUNDS PER ACRE AND INDEX-NUMBER OF YIELD OF *AMAN* RICE (NOT IN HUSK) FOR CIRCULAR CUTS OF DIFFERENT SIZES

serial no.	name of district	no. of cuts	yield in lbs. per acre			index-number of yield with yield based on largest size of cuts as 100	
			12.57 sq. ft.	50.27 sq. ft.	100.88 sq. ft.	12.57 sq. ft.	50.27 sq. ft.
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	Bakarganj	180	961.1	880.5	891.2	107.8	98.8
2	Bankura	109	750.4	686.3	723.3	103.8	94.9
3	Birbhum	87	947.1	864.0	883.8	107.2	97.8
4	Bogra	76	728.2	639.4	637.7	114.2	100.3
5	Burdwan	122	801.5	724.9	723.3	110.8	100.2
6	Chittagong	25	1,156.1	935.6	924.9	125.0	101.1
7	Dacca	113	922.4	734.0	758.7	121.6	96.7
8	Dinajpur	218	933.1	819.6	822.0	113.5	99.7
9	Faridpur	36	1,036.0	655.8	597.4	173.4	109.8
10	Hoogly	45	706.8	613.0	616.3	114.7	99.5
11	Howrah	38	604.8	420.5	401.6	150.6	104.7
12	Jalpaiguri	60	655.8	578.5	579.3	113.2	99.9
13	Jessore	86	723.3	613.8	621.3	116.4	98.8
14	Khulna	91	832.7	741.4	750.4	111.0	98.8
15	Malda	73	799.8	738.9	749.6	106.7	98.6
16	Midnapur	249	767.7	703.5	700.2	109.6	100.5
17	Murshidabad	55	870.6	785.8	777.6	112.0	101.1
18	Mymensingh	263	766.9	671.4	679.7	112.8	98.8
19	Nadia	38	680.5	598.2	613.8	110.9	97.4
20	Noakhali	58	789.9	701.1	669.0	118.1	104.8
21	Pabna	55	512.6	431.2	429.5	119.3	100.4
22	Rajshahi	83	805.6	687.1	665.7	121.0	103.2
23	Rangpur	204	794.1	715.9	715.9	110.9	100.0
24	Tipperah	92	704.4	598.2	610.6	115.4	98.0
25	24–Parganas	113	706.0	602.3	618.0	114.2	97.5
	total	2,569	812.2	704.4	706.8	—	—
	percentage	—	114.9	99.7	100	117.3	100.1
				weighted		unweighted	

At that time the idea I had in mind was that the unconscious pulling in of bordering plants was probably effective so long as the linear dimensions of the sample cuts were smaller than the human span.[1] In 1944 Jitendra Mohan Sen Gupta of the Statistical Laboratory suggested the use of circular-shaped cuts to be harvested with the help of an arm rotating over a pivot with a light stylus attached at the end of the rotating arm to catch the plants. This proved very convenient in practice, and various improvements were made. The latest model is now standard implement in our work. It is the usual practice at the present time to take three (or four) concentric circular cuts on the same field. This is done by drawing out the rotating (radius) arm to three (or four) fixed lengths one after another, and harvesting the appropriate cuts. In this way yield rates are obtained for three or four different sizes.

Illustrative results of certain experiments are given in Table 11. Two things are clear from the above table. Circular cuts of 12.57 sq. ft. are still over-estimating when the work is done on an extensive scale by a large field staff scattered over the whole province of Bengal, comprising about 60,000 or 70,000 square miles. Secondly, it is equally clear that there is practically no bias when the size of the cut is increased to 50.27 sq. ft. which is in good agreement with previous observations.

Along with extensive experiments carried out by the ordinary field staff, arrangements were also made to repeat the work at a few selected centres by experienced investigators working under the direct supervision of trained statisticians. Illustrative results are given below in Table 12, which shows the yield in lbs. per acre of *aman* (winter) rice at three centres in Bengal in 1945-46. The numerical values show that the systematic over-estimation with cuts of 2 ft. radius is no longer appreciable. This is confirmed by Student's paired "*t*" values of which are given at the bottom of the table.

TABLE 12. BENGAL CROP SURVEY, *AMAN* (WINTER) RICE, 1945–46: YIELD IN POUNDS PER ACRE FOR DIFFERENT SIZES OF CUTS

symbol	size of sample cut in sq. feet	mean yield in lbs. per acre with s.e.		
		Katwa ($n = 124$)	Gouripur ($n = 64$)	Sainthia ($n = 48$)
x_1	$\pi(2')^2 = 12.57$	1,261 ± 33	726 ± 25	1,593 ± 46
x_2	$\pi(4')^2 = 50.27$	1,283 ± 26	747 ± 20	1,518 ± 31
x_3	$\pi(5\frac{5}{8}')^2 = 100.88$	1,304 ± 26	761 ± 19	1,523 ± 28
x_4	$\pi(8')^2 = 201.06$	1,340 ± 29	764 ± 18	1,495 ± 24
		Values of Student's paired "*t*"		
$x_1 - x_2$	—	1.53	1.33	2.18 *
$x_2 - x_3$	—	2.55 *	1.81	0.43
$x_3 - x_4$	—	3.07 †	0.63	2.47 *

* Significant at 5 per cent level. † Significant at 1 per cent level.

---

[1] Dr. P. V. Sukhatme in a letter to *Nature* of May 11th, 1946, has published some material showing the over-estimation of crop yields with cuts of a small size. He writes: "The reason for over-estimation appears to be the human tendency to include border plants inside the plot. This factor becomes serious when the perimeter of the plot is large in proportion to its area." This is exactly what I had suggested in the paper written in 1942. Dr. Sukhatme has not referred to my paper, or is not aware of the observations made by me much earlier. I have recently sent a letter to *Nature* on this subject.

We thus find that when the work is done under adequate statistical supervision, the over-estimation with cuts of a small size becomes practically negligible. In view of this finding, I believe that unconscious pulling in of bordering plants may not be the only, or even the most important factor in this matter. In fact, I am now inclined to adopt a line of explanation offered by F. Yates in his paper on bias in sampling in the *Annals of Eugenics* in 1936. Discussing the observed bias in crop yields harvested from within hoops (of area 10 sq. ft.) supposed to have been thrown at random on fields in the United Provinces, Yates suggested:

> "The bulk of the bias, however, is probably due to the tendency, conscious or unconscious, to cast the hoop on the good parts of the crop."

It is likely that there are patches of greater fertility distributed either in a random manner or in a mildly patterned fashion over each field. In locating the sample cut, the investigator may unconsciously tend to favour such more fertile patches by shifting about the "random point" to some extent. From the fact that the bias appears to become negligible beyond 40 or 50 sq. ft., it would seem, without entering into detailed investigations, that the average size of such patches would be probably fairly small and of the order of only a few square feet. Unconscious pulling in of bordering plants may also be a contributing factor.

Another point is perhaps just worth mentioning. In case there is any bias in favour of locating the sample cuts on or near more fertile patches, then one would expect over-estimation with the smaller cuts to be followed by compensating under-estimation in cuts of intermediate size. Some of the figures of Table 10 may perhaps be indicative of this, but more precise and detailed experimentation is necessary to settle this point.

A certain amount of loss of grain in the actual process of harvesting, threshing, drying, weighing, etc., is inevitable which will necessarily introduce a component of "error". The proportional error arising from this factor may be independent of or may increase or decrease with the size of the cut. In case the loss of grain is proportionally smaller with small cuts (owing to the greater care with which the different processes can be carried out), then there might be an apparent over-estimation with small cuts, which, however, would really indicate an under-estimation with large cuts.

It is pertinent to observe in the present connection the danger of under-estimation in using cuts of a comparatively large size which are demarcated by pegs and ropes on the field. It is doubtful whether ordinary investigators can make accurate measurements of distance on the field. Any sagging of the ropes is bound to reduce the real size of the cut, and hence lead to under-estimation. There are serious difficulties in deciding the allowance to be made for the boundary ridges or demarcations between adjoining fields, a point which is of particular importance in a province like Bengal, where the average size of the field is very small and less than half an acre, so that the perimeter is large in comparison with the area of the field. In fact, it is difficult to define the "whole field" in an unambiguous manner; and until this is done the concept of total production must also remain to some extent vague. This subject clearly requires much further study.

In the meantime I may perhaps mention that the over-estimation of cuts of a very small size can itself be used as an excellent control in the following way. As already mentioned, yield rates are being obtained by the Statistical Institute for three (or four) concentric circular cuts of different sizes on the same plot. If the work is done honestly by ordinary investigators, then yield rates based on sample cuts of the smallest size would show over-estimation in comparison with results based on sample cuts of larger size. Results based on the two (or three) larger cuts should, however, be in good agreement. At the same time, the variance of the yield per acre should decrease with increasing sizes of sample cuts. Besides being extremely convenient and economical, concentric circular cuts thus enable a two-fold control (one over mean rates of yield and the other over variances) to be exercised at the point of collection of the primary material.

BENGAL LABOUR ENQUIRY: JAGADDAL 1941, 1942 AND 1945

As already mentioned, in socio-economic surveys special efforts have been made to study the personal equation or bias of the investigating staff. Here also the use of replicated sub-samples proved to be of great value. I shall give illustrative examples from certain enquiries carried out in 1941, 1942 and 1945 in an industrial area at Jagaddal, about 20 miles to the north of Calcutta. The Government of Bengal had asked us to survey family budgets, housing, and other economic conditions of factory workers in this area.

Design of the survey. The general approach can be best studied from the design of the enquiry of 1941. In order to ensure proper randomization, the addresses of working-class employees were obtained from the factories located in this area, on the basis of which a preliminary survey was made of the geographical distribution of working-class families at Jagaddal, which is a municipal area. The number of families residing in different groups of buildings was noted on large-scale maps. A sample check was also carried out, but no attempt was made to attain high precision in the preliminary survey, as this was not considered to be of great importance.

On the basis of the preliminary survey the area was divided into five blocks (which were of unequal size and irregular in shape and), which were demarcated on certain general considerations. For example, blocks nos. 1, 2 and 3 were located in the immediate vicinity of factories; block no. 4 was within the town, but was situated at a greater distance from the factories themselves; finally, block no. 5 covered the area outside municipal limits. For present purposes the first three blocks were most important; block no. 4 of moderate importance, while block no. 5 was included more or less as a control and for comparative purposes. The buildings from which the families were to be surveyed were picked out purely at random on maps.

The sample units within each block were divided into 5 equal sub-samples, each of which was independently random and covered the whole of the block. These five sub-samples thus constituted five independent networks of sample units within each block. Each such sub-sample was assigned to a different investigator. In this way information for each block was collected by five investigators, all of whom in consequence worked in all the blocks. The number of families surveyed in 1941 and 1942 by the different investigators in different blocks are shown in Table 13.

Possibilities of analysis of variance. The interesting point is that the design of the survey makes it possible to carry out a Fisherian analysis of variance. For example, the 1941 and 1942 designs resemble in appearance Fisherian Latin squares. These are, however, not true Latin squares, but only analogues, as the sub-samples are not geographically distinct. In fact there is no intrinsic difference between the Latin square and the

TABLE 13. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941–42: NUMBER OF FAMILIES SURVEYED BY BLOCKS AND INVESTIGATORS

blocks	investigators					total
	1	2	3	4	5	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
			1941 survey			
1	33	33	33	31	27	157
2	36	36	31	31	34	168
3	37	36	45	34	29	181
4	20	18	20	23	18	99
5	7	9	8	6	7	37
total	133	132	137	125	115	642
			1942 survey			
1	41	37	40	39	38	195
2	41	36	46	39	35	197
3	38	42	39	41	37	197
4	19	24	19	24	21	107
5	9	9	8	8	10	44
total	148	148	152	151	141	740

randomized block, so that an analysis of variance can be carried out for any two or more or all the blocks at the same time.

The unequal numbers in the different block-investigator cells introduce a difficulty. There are three possibilities. If the first three blocks are selected the numbers are not very different, so that a simple analysis of variance on orthodox lines would give approximate results. A second possibility (for the same three blocks) is to equalize the number in each cell by rejecting an appropriate number of schedules at random from each cell. Finally, the method of fitting constants can of course be used, but would involve a great deal of computational labour.

The results of an analysis of variance (for the first three blocks) for the family budget enquiry at Jagaddal in 1942 are given for a few selected items in Table 14. The total number of household schedules was 589 with 588 degrees of freedom, which can be split up into two degrees of freedom for comparisons between the three blocks, four degrees of freedom for comparisons between five investigators, and eight degrees of freedom for the interaction making up altogether fourteen degrees of freedom for comparisons between sub-samples.

In the standard Fisherian analysis of variance it is usual to use the interaction as error to test the significance of difference between blocks, etc. In the present case a direct estimate of the error can be made from the variations "within block-investigators cells" (that is, within each portion of the sample for which the material is collected by the same investigator in the same geographical area) based on 574 degrees of freedom.

It is also possible to carry out a similar analysis after equalizing the frequencies in each cell by rejecting an appropriate number of schedules at random. This was done, and the results are given in Table 14 for convenience of comparison. The total number of schedules used for this purpose was 525, so that the total variance was based on 524 degrees of freedom. Deducting fourteen degrees of freedom for comparisons among the sub-samples, this left 510 degrees of freedom for estimating the variance within sub-samples.

The ratios of variance with unequal as well as equalized frequencies in each cell are shown in the lower half of Table 14. The results are broadly similar, but equalization of frequencies makes the comparison more stringent, as a number of ratios are significant here, but not in the analysis with unequal frequencies.

TABLE 14. BENGAL LABOUR ENQUIRY—JAGADDAL, 1942: ANALYSIS OF VARIANCE

(3 blocks × 5 investigators)

sources of variation	d.f.	age in years	expenditure in rupees per month per capita			consumption of cereals in lbs. per head per month
			total	food	cereals	
(1)	(2)	(3)	(4)	(5)	(6)	(7)
values of variance						
(*a*) actual (unequal) number of families per cell ($n = 589$)						
between blocks	2	76.1	606.5	23.4	0.08	82.5
between investigators	4	211.6	220.4	19.5	1.3	61.8
blocks × investigators	8	134.1	114.1	11.2	1.9	212.4
between sub-samples	14	147.9	214.8	15.3	1.5	151.1
within sub-samples	574	112.7	172.9	10.2	0.6	97.3
total	588	113.5	173.9	10.3	0.6	98.6
mean value	—	38.84	19.16	7.80	3.05	35.59
standard deviation	—	9.71	11.41	3.15	0.73	8.64
coefficient of variation	—	25.0	59.5	40.4	23.9	23.8
(*b*) equalized number of families per cell ($n = 525$)						
between blocks	2	62.13	805.52	36.5	0.07	79.6
between investigators	4	304.84	275.78	22.3	1.25	114.7
blocks × investigators	8	78.47	129.38	9.6	1.47	152.8
between sub-samples	14	140.81	267.80	16.8	1.21	132.0
within sub-samples	510	127.74	168.12	9.9	0.49	100.3
total	524	128.09	170.78	10.1	0.51	100.8
mean value	—	39.31	20.86	8.15	3.09	35.96
standard deviation	—	11.32	13.07	3.18	0.71	10.04
coefficient of variation	—	28.7	62.6	39.0	22.9	27.9
ratios of variance						
(*a*) actual (unequal) number of families per cell ($n = 589$)						
between blocks	2	0.68	3.51*	2.29	0.13	0.85
between investigators	4	1.88	1.27	1.91	2.17	0.63
blocks × investigators	8	1.19	0.66	1.10	3.17†	2.18
between sub-samples	14	1.31	1.24	1.50	2.50†	1.55
within sub-samples	574	—	—	—	—	—
total	588	—	—	—	—	—
(*b*) equalized number of families per cell ($n = 525$)						
between blocks	2	0.49	4.79†	3.69*	0.14	0.80
between investigators	4	2.39*	1.64	2.26	2.55*	1.15
blocks × investigators	8	0.61	0.77	0.98	3.00†	1.53
between sub-samples	14	1.10	1.59	1.70	2.47†	1.32
within sub-samples	510	—	—	—	—	—
total	524	—	—	—	—	—

* Significant at 5 per cent level. † Significant at 1 per cent level.

Several points are worth noting. Firstly, it will be noticed that, except in the case of monthly expenditure on cereals, the interaction (blocks × investigators) variances are not significant, showing that the Fisherian assumption of the interaction variance being about the same as the "true" within-cell variance is broadly confirmed.

The significant value itself was further analysed, block by block and investigator by investigator, by Janardan Poti and K. C. Cheriyan; and it was found that the abnormally high values were due to one single investigator in one particular block. This shows the possibilities of deeper analysis offered by replicated inter-penetrating samples.

Investigator differences were insignificant in two cases, and significant at 5 per cent level in two cases. Personal equations had not been completely eliminated but their influence was not large.

Nagpur Labour Enquiry 1942–43. I am giving a second example from a labour enquiry carried out in Nagpur in 1942–43 by M. P. Shrivastava, at that time a worker of the Institute, under the technical guidance of the Indian Statistical Institute. In this case the design was arranged in the form of a randomized block of five zones and four investigators with practically the same number of family schedules—namely, fifty in each block-investigator cell. The analysis of variance is shown in Table 15.

TABLE 15. NAGPUR FAMILY BUDGET ENQUIRY, 1943 : ANALYSIS OF VARIANCE
(5 zones × 4 investigators)

sources of variation	d.f.	total income	monthly expenditure		
			total	food	cereals
(1)	(2)	(3)	(4)	(5)	(6)
		(a) values of variance			
between zones	4	4,439.58	3,707.91	708.41	206.83
between investigators	3	85.43	597.08	77.09	3.70
zones × investigators	12	382.54	397.28	177.75	49.80
between sub-samples	19	1,189.74	1,127.07	237.58	75.61
within sub-samples	977	401.57	384.71	84.73	24.99
total	996	424.67	398.87	88.33	25.95
mean values (in rupees)	—	36.09	34.96	20.09	11.41
standard deviations	—	18.60	16.12	8.53	4.96
coefficients of variation	—	51.5	46.1	42.5	43.5
		(b) ratios of variance			
between zones	4	11.06†	9.64†	8.36†	8.28†
between investigators	3	0.21	1.55	0.91	0.15
zones × investigators	12	0.95	1.03	2.10*	2.00*
between sub-samples	19	2.96†	2.93†	2.80†	3.02†
within sub-samples	977	—	—	—	—
total	996	—	—	—	—

* Significant at 5 per cent level. † Significant at 1 per cent level.

It will be noticed that in this case the sub-samples are all significantly differentiated. Block differences are also highly significant in each case.

It may be mentioned here that in the Nagpur enquiry the blocks had been deliberately demarcated on lines which could almost certainly be expected to show large differences in income and conditions of living. For example, one block consisted almost exclusively of employees of the Bengal Nagpur Railway Workshop, who certainly earned higher wages and had generally a higher standard of living. The other blocks also were differentiated in many ways.

It is, however, most satisfactory to find that investigator's differences were negligible in every case, showing that personal equations had been completely eliminated. It is only by an analysis of the present kind that one can be certain on this point.

Margin of error of cost of living index : Similar studies are also possible in regard to the margin of error of the index-number of cost of living. For example, the C.L. index-number can be calculated separately for each block-investigator cell. Adopting 1941 consumption pattern as base and 1942 and 1945 prices, the results are given in Table 16. In this table the number of 1941 family schedules on which each index-number is based is given within brackets. From the values of the index in different cells it is possible to calculate the standard error of the marginal index-numbers separately for each block or for each investigator; and finally, in the same way, the standard error of the over-all index number. These are shown in Table 16.

TABLE 16. BENGAL LABOUR ENQUIRY: JAGADDAL COST-OF-LIVING INDEX 1942 AND 1945, WITH 1941 AS BASE BY BLOCKS AND INVESTIGATORS (WITH SIZE OF SAMPLE WITHIN BRACKETS)

blocks	investigators					total	
	1	2	3	4	5	size of sample in '41	mean index with s.e.
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			1942 cost-of-living index				
1	120 (33)	122 (33)	122 (33)	124 (31)	124 (27)	157	122 ± 0.67
2	122 (36)	120 (36)	121 (31)	124 (31)	118 (34)	168	121 ± 1.88
3	121 (37)	123 (36)	121 (45)	121 (34)	121 (29)	181	121 ± 0.36
4	119 (20)	122 (18)	120 (20)	123 (23)	121 (18)	99	121 ± 0.65
5	121 (7)	124 (9)	124 (8)	124 (6)	124 (7)	37	123 ± 0.53
n (1941)	(133)	(132)	(137)	(125)	(115)	642	—
mean index	121 ± 0.46	122 ± 0.57	121 ± 0.41	123 ± 0.41	121 ± 1.04	—	122 ± 0.33
			1945 cost-of-living index				
1	270 (33)	280 (33)	275 (33)	271 (31)	279 (27)	157	276 ± 2.0
2	263 (36)	265 (36)	284 (31)	269 (31)	267 (34)	168	269 ± 3.6
3	264 (37)	274 (36)	278 (45)	272 (34)	269 (29)	181	272 ± 2.4
4	263 (20)	274 (18)	271 (20)	297 (23)	263 (18)	99	275 ± 6.2
5	260 (7)	279 (9)	296 (8)	274 (6)	284 (7)	37	279 ± 6.2
n (1941)	(133)	(132)	(137)	(125)	(115)	642	—
mean index	265 ± 1.38	273 ± 2.56	279 ± 2.69	276 ± 4.57	271 ± 2.87	—	273 ± 1.64

It will be noticed that the 1942 mean index is 122 with a standard error of about 0.4 (which is only about 0.3 per cent of the mean value). The standard errors of marginal index-numbers are quite small and usually less than 1.

The 1945 C.L. index has a standard error of 1.6 or about 0.5 per cent of the mean index-number 273. Standard errors in 1945 were, however, much higher in comparison with standard errors in 1942. This is partly accounted for by the fact that whereas price relatives had varied very little among themselves (roughly about 10 per cent of their mean values) in 1942, price relatives had varied among themselves about 80 per cent of their mean values in 1945.

Analysis of variance has been also carried out both for five blocks × five investigators and for three blocks × five investigators (for the three blocks in which the number of schedules are nearly equal). The results are shown in Table 17.

TABLE 17. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945 : ANALYSIS OF VARIANCE : COST-OF-LIVING INDEX FOR 1942 AND 1945, WITH 1941 AS BASE

source of variation	5 blocks × 5 investigators			3 blocks × 5 investigators			expected ratio	
	d.f.	variance	ratio	d.f.	variance	ratio	5%	1%
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			1942 with 1941 as base					
between blocks	4	5.25	2.40	2	2.60	0.80	4.46	8.65
between investigators	4	4.50	2.12	4	2.06	0.63	3.84	7.01
residual	16	2.12	—	8	3.27	—	—	—
total	24	3.04	—	14	2.83	—	—	—
			1945 with 1941 as base					
between blocks	4	59.75	0.81	2	37.80	1.54	4.46	8.65
between investigators	4	194.00	2.64	4	69.00	2.81	3.84	7.01
residual	16	73.44	—	8	24.55	—	—	—
total	24	91.25	—	14	39.14	—	—	—

It is seen that block and investigator differences were both insignificant for both 1942 and 1945. This is satisfactory and gives confidence in using the present index-numbers.

It is of course also possible to study the cost-of-living index for different groups of items separately in each block-investigator cell. Illustrative results are given in Table 18.

There are many other possibilities of breakdowns. For example, the cost-of-living index can be studied by groups of items for different over-all size of households. Relative figures are given in Table 19, from which it would be noticed that the change in the cost-of-living index-number did not depend appreciably on the size of the household.

Similar detailed figures for the change in the cost of living in 1942 and 1945 for different families in different expenditure levels are given in Table 20. Here also the cost of living appears to have changed more or less in the same way in all expenditure levels. It must be remembered, however, that the enquiry was restricted to working-class families in which the consumption pattern was fairly homogeneous.

TABLE 18. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945 : COST-OF-LIVING INDEX BY BLOCK × INVESTIGATOR

block no.	investigator no.	food		clothing		fuel and light		miscellaneous		total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
I	1	120	276	159	310	126	258	114	275	120	270
	2	121	279	159	304	129	285	117	293	122	280
	3	122	277	160	310	127	277	119	285	122	275
	4	125	280	156	301	130	289	120	271	124	271
	5	120	280	160	314	125	275	125	287	124	279
	total	122	280	159	308	127	276	119	283	122	275
II	1	123	273	159	306	127	280	114	253	122	263
	2	121	270	158	304	126	289	114	262	120	265
	3	122	279	155	297	125	267	117	303	121	284
	4	125	272	156	305	127	283	120	271	124	269
	5	121	270	156	304	126	281	113	273	118	267
	total	122	273	157	304	126	280	116	273	121	269
III	1	122	275	156	297	129	293	114	249	121	264
	2	127	277	161	304	124	281	114	279	123	274
	3	121	280	160	306	125	288	117	289	121	278
	4	121	281	153	298	125	283	115	266	121	272
	5	122	269	159	305	125	280	115	278	121	269
	total	123	277	158	302	126	286	115	272	121	272
IV	1	119	272	153	277	126	280	114	256	119	263
	2	124	274	158	289	126	285	115	292	122	274
	3	121	274	157	242	131	303	114	290	120	271
	4	122	306	161	313	124	273	120	311	123	297
	5	124	275	154	302	123	274	112	253	121	263
	total	122	281	157	293	126	283	115	283	121	275
V	1	120	266	159	304	128	259	118	281	121	260
	2	124	288	158	303	127	266	119	292	124	279
	3	119	273	162	312	125	266	123	340	124	296
	4	123	283	158	315	125	240	120	280	124	274
	5	125	280	163	309	131	276	114	315	124	284
	total	122	277	160	308	127	264	119	308	123	279
I–V	1	121	274	157	303	127	278	114	258	121	265
	2	123	276	159	304	126	284	115	280	122	273
	3	121	278	159	296	126	282	118	293	121	279
	4	124	283	157	305	127	282	119	277	123	276
	5	123	276	158	307	126	278	117	278	121	271
	total	122	277	158	303	126	280	117	276	122	273

N.B.—No change in "rent and tax."

TABLE 19. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945: COST-OF-LIVING INDEX BY SIZE OF FAMILY

size of family in 1941	no. of families in 1941	food		clothing		fuel and light		miscellaneous		total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
1	266	122	274	160	310	124	274	117	272	121	270
2	94	124	275	156	298	127	286	114	285	123	273
3	69	123	280	157	299	127	287	114	275	122	272
4	70	122	286	156	297	129	288	117	281	122	277
5	56	121	285	154	296	127	284	118	287	122	279
6	34	122	275	157	297	129	282	118	285	122	271
7	19	124	280	155	295	131	284	120	286	125	275
8	13	122	280	157	302	132	294	126	284	124	274
9	12	123	285	154	294	131	273	120	292	124	278
10	1	125	294	160	324	119	228	119	272	123	270
11	3	124	288	159	300	127	304	124	266	127	279
13	3	125	279	157	296	138	316	128	286	128	278
14	1	124	284	161	306	139	323	119	346	126	236
total	641	122	277	158	305	126	280	117	276	122	273

N.B.—No change in "rent and tax."

TABLE 20. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945 : COST-OF-LIVING INDEX BY EXPENDITURE LEVELS IN RUPEES PER MONTH PER FAMILY

expenditure levels in 1941	no. of families in 1941	food		clothing		fuel and light		miscellaneous		total	
		1942	1945	1942	1945	1942	1945	1942	1945	1942	1945
0– 9	22	123	276	156	308	126	285	121	301	123	282
10–19	189	122	275	158	306	126	201	114	281	121	268
20–29	199	122	275	158	306	125	278	115	273	121	270
30–39	107	123	277	157	301	126	279	117	268	122	269
40–49	56	122	291	157	301	128	285	118	279	122	281
50–59	30	122	281	161	305	129	286	123	276	124	273
60 & above	38	124	276	161	304	129	284	123	277	125	272
total	641	122	277	158	305	126	280	117	276	122	273

N.B.—No change in "rent and tax."

Comparison of consumption in 1941 *and* 1945. As already mentioned, family budget enquiries were repeated in the same area in 1941 and also in 1942 and again in 1945. This enables a direct comparison being made of levels of consumption in different years. Somewhat similar material also happens to be available for Calcutta middle-class families in 1939 and in 1945. As prices and consumption patterns had not changed very much in Bengal between 1939 and 1941, the above two series can also be used to compare (at least approximately) the changes in patterns of consumption in working-class and middle-class families.

Relevant data are given in Table 21. The name of the item is given in col. 1, and the unit of measurement in col. 2. The over-all *per capita* consumptions in 1945 and 1941

TABLE 21. COMPARISON OF CONSUMPTION IN DIFFERENT YEARS FOR WORKING AND MIDDLE-CLASS FAMILIES IN BENGAL

item	unit per head per year	Jagaddal working-class				Calcutta middle-classes			
		consumption *per capita* per year		index-number, 1945 : 1941		consumption *per capita* per year		index-number, 1945 : 1939	
		1945	1941	con-sump-tion	prices	1945	1939	con-sump-tion	prices
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
rice and products	lbs.	214.0	321.8	66.5	273	221.4	224.7	98.5	308
wheat and products	,,	133.3	113.6	117.4	187	85.6	74.1	115.5	258
all cereals	,,	345.6	435.3	79.4	—	306.9	298.7	102.8	—
pulses	,,	68.3	70.0	97.6	366	34.5	32.1	107.6	317
potatoes	,,	50.2	64.8	77.5	533	54.9	83.1	66.0	377
milk and products	,,	18.3	43.2	42.3	300	77.8	174.5	44.6	333
butter and ghee	,,	0.66	3.9	16.8	210	1.6	16.5	10.0	333
oil	,,	20.5	15.2	134.7	304	19.1	23.0	83.0	235
salt	,,	12.0	15.9	75.6	190	11.8	17.3	68.3	300
sugar and gur	,,	26.4	15.0	176.0	185	27.9	48.6	57.5	172
fish	,,	12.1	13.1	92.2	290	24.5	50.2	48.8	589
meat	,,	8.0	12.0	66.6	343	7.7	11.5	67.0	360
eggs	no.	—	—	—	—	11.8	42.4	28.0	562
coal	lbs.	297.1	279.0	106.5	377	—	—	—	—
fire wood	,,	302.0	405.7	74.4	330	—	—	—	—
cloth (dhuti)	yds.	6.9	9.9	69.7	283	—	—	—	—
clothing (lungi)	,,	0.6	0.5	120.0	416	—	—	—	—
number of families	—	755	641	—	—	610	1,151	—	—
,, persons	—	2,313	1,866	—	—	4,335	10,539	—	—
,, persons per family	—	3.06	2.91	—	—	7.11	9.16	—	—

for working-class families are given in cols. 3 and 4 respectively. The consumption in 1945 expressed as a percentage of the corresponding consumption in 1941 is given in col. 5; and the price ratios in col. 6. Similar data for *per capita* consumptions in 1945 and 1939 in middle-class families are given in cols. 7 and 8. The ratio of consumptions is given in col. 9, and price ratios in col. 10.

I should mention here that food began to be rationed in 1943. The ration of rice was somewhat restricted owing to shorter supply, and the price control was arranged in such a way as to make "wheat" and "wheat products" comparatively cheaper. Sugar was strictly rationed, and there was a partial rationing of oil.

The consumption of all cereals had dropped very appreciably to about 79 per cent in working-class families, although there was an increase in the consumption of wheat, owing no doubt to more favourable prices. In middle-class families the cereal consumption (which had been originally much lower in comparison with working-class families) had remained more or less steady, but there was a higher consumption of wheat. The consumption of sugar had increased a great deal (to 176 per cent) among working-class families, but had decreased appreciably in middle-class families, due almost certainly to rationing. The consumption of oil had increased somewhat in working-class families, which was probably due to partial rationing and as an offset against the practical discontinuance of the use of higher-quality fats like butter or ghee (clarified butter). In practically all other items the consumption had decreased very seriously in 1945 in both working-class and middle-class families at Jagaddal and in Calcutta.

Margin of error of index-number of earnings and consumption. As information had been collected about total earnings of different families in 1941 and 1945, it is possible to calculate the average increase in earnings. The geographical blocks had been kept the same in both years. It is possible therefore to calculate average earnings for each block separately in 1941 and 1945, and hence to calculate the ratio of earnings for each individual block. Similar material is also available for a number of other items, like monthly expenditure, consumption of cereals, pulses, vegetables, vegetable oil, meat and fish, dairy products, and clothing. Index-numbers for 1945 with 1941 as base can be therefore calculated for all these items for each of the five geographical blocks at Jagaddal. These index-numbers are shown in Table 22. Similar figures were available for index-numbers for 1942 with 1941 as base, but these have not been printed here.

Table 22 supplies some idea of variations in the index-numbers from block to block. It is possible to form an unweighted average index for the whole area, and also to calculate the unweighted standard error from the figures for the different blocks. These unweighted average index-numbers for 1942 on 1941 are shown in col. 3, and for 1945 on 1941 in col. 4 of Table 24. Information regarding the size of the samples in 1941 and in 1945 in the different blocks is given together with their harmonic means at the top of each column in Table 22. Using such harmonic means as weights, a series of weighted averages with standard errors were calculated, and are shown in col. 5 of Table 24.

Mean values of monthly earnings, expenditure, and consumption of various commodities for working-class families in Jagaddal in 1941, 1942 and 1945 are shown in Table 23. In this table figures have been given both on a "per family" and "*per capita*"

TABLE 22. BENGAL LABOUR ENQUIRY—JAGADDAL, 1945 AND 1941 : INDEX-NUMBERS FOR 1945 WITH 1941 AS BASE (= 100)

	(*a*) per family basis					(*b*) *per capita* basis				
	block 1	block 2	block 3	block 4	block 5	block 1	block 2	block 3	block 4	block 5
n (1941)	157	168	181	99	37	516	484	434	268	164
n (1945)	195	212	192	104	51	580	601	467	374	284
items harmonic mean	174	188	186	101	43	546	536	450	312	208
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
cost of living	275	269	272	275	279	275	269	272	275	279
earnings	173	195	207	228	221	191	199	204	172	176
expenditure	165	179	194	212	206	182	183	191	160	164
cereals	78	80	80	107	81	86	82	79	81	65
pulses	89	115	93	126	98	98	118	91	95	78
vegetables	90	98	103	113	92	99	100	101	85	73
vegetable oil	94	149	176	166	136	104	153	173	111	108
meat and fish	71	78	104	95	72	79	80	102	72	57
species and salt	75	74	71	88	109	83	76	70	67	87
dairy products	30	48	46	47	32	34	50	46	36	26
clothing	73	73	83	87	92	81	75	81	66	73

TABLE 23. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945 : AVERAGE EARNING AND EXPENDITURE AND QUANTITY CONSUMED PER MONTH

serial no.	item	unit of measurement per month	1941	1942	1945
(1)	(2)	(3)	(4)	(5)	(6)
		average size of family	2.91	2.79	3.06
		(*a*) per family basis			
1.	earning	rupees	26.93 ± 0.92	36.38 ± 1.22	55.40 ± 5.16
2.	expenditure	”	30.11 ± 1.16	35.57 ± 1.15	57.61 ± 5.38
3.	cereals	pounds	103.41 ± 5.70	98.08 ± 5.53	87.49 ± 7.63
4.	pulses	”	15.47 ± 0.70	12.61 ± 0.56	15.93 ± 0.82
5.	vegetables	”	32.32 ± 2.20	31.41 ± 1.52	31.56 ± 2.55
6.	vegetable oil	”	3.56 ± 0.19	3.48 ± 0.19	5.10 ± 0.47
7.	meat and fish	”	6.09 ± 0.35	5.90 ± 0.33	5.00 ± 0.33
8.	species and salt	”	5.74 ± 0.33	5.74 ± 0.29	4.96 ± 0.97
9.	dairy products	”	13.31 ± 1.69	12.08 ± 1.46	5.10 ± 0.88
10.	clothing	yds. per year	60.49 ± 2.11	55.42 ± 2.52	49.77 ± 5.78
		(*b*) *per capita* basis			
1.	earning	rupees	9.01 ± 0.42	13.32 ± 0.53	16.54 ± 1.09
2.	expenditure	”	9.93 ± 0.35	12.08 ± 0.55	17.20 ± 1.15
3.	cereals	pounds	33.37 ± 0.74	31.76 ± 0.53	26.08 ± 1.48
4.	pulses	”	5.16 ± 0.25	4.20 ± 0.16	4.98 ± 0.64
5.	vegetables	”	10.45 ± 0.45	10.55 ± 0.56	9.46 ± 0.62
6.	vegetable oil	”	1.15 ± 0.04	1.13 ± 0.02	1.52 ± 0.21
7.	meat and fish	”	1.97 ± 0.02	2.04 ± 0.10	1.52 ± 0.14
8.	species and salt	”	1.85 ± 0.06	1.89 ± 0.06	1.40 ± 0.06
9.	dairy products	”	4.05 ± 0.23	3.79 ± 0.33	1.50 ± 0.06
10.	clothing	yds. per year	19.92 ± 0.64	18.33 ± 0.58	14.70 ± 0.87

TABLE 24. BENGAL LABOUR ENQUIRY—JAGADDAL, 1941, 1942 AND 1945: INDEX-NUMBERS OF EARNINGS, EXPENDITURE AND CONSUMPTION

serial no.	item	1942 : 1941	1945 : 1941		
		unweighted average of ratios	unweighted average of ratios	weighted average of ratios	ratios of average
(1)	(2)	(3)	(4)	(5)	(6)
	cost-of-living index	122 ± 0.33	273 ± 1.64	272.8 ± 1.3	—
	(*a*) index-number on "per family" basis				
1.	earnings	137 ± 4.3	205 ± 9.8	199.1 ± 8.3	205.8 ± 20.50
2.	expenditure	121 ± 5.1	191 ± 8.6	186.0 ± 7.3	191.3 ± 19.35
3.	cereals	98 ± 6.1	85 ± 5.5	83.5 ± 4.4	84.6 ± 8.73
4.	pulses	86 ± 5.4	104 ± 7.0	103.1 ± 6.2	103.0 ± 7.09
5.	vegetables	105 ± 6.9	99 ± 4.1	99.2 ± 3.4	97.6 ± 10.33
6.	vegetable oil	101 ± 5.5	144 ± 4.3	144.1 ± 14.0	143.3 ± 15.28
7.	meat and fish	106 ± 7.0	84 ± 6.6	85.3 ± 6.1	82.1 ± 7.20
8.	species and salt	105 ± 7.0	83 ± 7.0	77.7 ± 4.3	86.4 ± 17.61
9.	dairy products	100 ± 9.3	41 ± 4.0	41.8 ± 3.5	38.3 ± 8.37
10.	clothing	92 ± 5.8	82 ± 3.8	78.9 ± 2.8	82.3 ± 9.98
	(*b*) index-number on "*per capita*" basis				
1.	earnings	141 ± 7.2	188 ± 6.3	191.5 ± 5.1	183.6 ± 14.83
2.	expenditure	122 ± 5.7	176 ± 6.0	179.1 ± 4.8	173.2 ± 13.09
3.	cereals	97 ± 3.0	79 ± 3.6	80.5 ± 3.4	78.2 ± 4.75
4.	pulses	85 ± 4.3	96 ± 6.5	99.2 ± 5.6	96.5 ± 13.22
5.	vegetables	106 ± 7.5	92 ± 5.5	94.9 ± 4.1	90.5 ± 7.15
6.	vegetable oil	98 ± 3.6	130 ± 14.0	133.4 ± 5.8	132.2 ± 18.52
7.	meat and fish	109 ± 8.1	78 ± 7.3	80.5 ± 6.5	77.2 ± 7.31
8.	species and salt	105 ± 5.0	77 ± 3.8	76.3 ± 3.0	75.7 ± 4.14
9.	dairy products	100 ± 9.0	38 ± 4.3	40.3 ± 4.2	37.0 ± 2.58
10.	clothing	93 ± 3.4	75 ± 2.8	76.4 ± 2.2	73.8 ± 4.95

basis, and the standard errors have been calculated from the block figures in 1945 and 1942. From this material, index-numbers can be prepared for 1945 with 1941 as base by dividing the mean values in 1945 by the corresponding mean values in 1941. One measure of the standard error of the index-number can also be obtained by compounding the standard error of the two mean values. These index-numbers are shown in col. 6 of Table 24, in which the C.L. index-number has been also given at the top for convenience of reference.

Index-numbers for 1945, calculated in three different ways, are given in cols. 4, 5 and 6 of Table 24. It will be noticed that it does not matter very much which particular method of calculation is followed, as the three index-numbers are in broad agreement.

It is of course possible to make a deeper analysis by taking into consideration the sampling errors within each block, which however, would entail a great deal of computational labour. My purpose here is to draw attention to the possibility of making rough but quick comparisons by using the mean values for replicated sub-samples.

I have no desire to discuss the economic implications, but may just note that, with a cost-of-living index of 273 and an index-number of earnings of 205, the index-number of real wages was 75 in 1945. The index-numbers of consumption shown in col. 5 of Table 21 or cols. 5 or 6 of Table 24 are in broad agreement with this value. Consumption of essential articles had naturally remained more steady, while consumption of other commodities had decreased more appreciably. I am not competent to discuss the economic aspects of the subject. My object is to point out that even approximate values of the margin of error are likely to be useful in obtaining a more critical appreciation of the economic situation.

I may mention here that in working-class family enquiries at Jagaddal arrangements had been made to keep a certain number of families common in the enquiries in different years. This enables a direct comparison of the change in the consumption pattern being made on a family or household basis. In fact the form in which the material has been collected makes it possible to undertake a critical study of such enquiries. Owing to lack of resources, we have not been able to take up this work seriously, but Ambika Ghosh and H. K. Chaturvedi are doing what they can.

TABLE 25. BENGAL SAMPLE SURVEY (1936–42) : ESTIMATED CONSUMPTION OF FOOD ITEMS IN POUNDS PER HEAD PER YEAR BY EXPENDITURE LEVELS

food items	levels of family expenditure in rupees per month								all levels
	0–10	10–20	20–30	30–50	50–100	100–200	200–300	300+	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
rice	310.22	371.93	370.28	360.41	342.31	278.95	257.55	262.49	341.48
rice products	4.11	5.76	10.70	9.87	10.70	10.70	9.87	5.76	8.23
wheat products	1.64	7.41	13.16	17.28	26.23	44.43	59.24	75.70	19.75
all cereals	315.97	385.10	394.14	387.56	379.34	334.08	326.66	343.95	369.46
pulses	15.63	27.98	27.03	36.20	32.91	31.27	34.56	36.20	32.91
salt	9.87	13.27	13.99	14.07	15.63	15.63	17.28	18.10	15.30
potatoes	12.34	26.33	32.09	37.03	48.55	63.36	79.81	108.61	48.55
vegetables	22.22	18.45	39.50	44.43	65.00	94.21	109.44	139.88	64.26
oil	5.76	9.87	11.52	12.34	15.63	19.75	21.39	25.51	14.81
sugar and gur	3.29	9.05	12.34	14.42	23.86	54.31	36.20	41.96	24.31
fish	6.58	8.23	11.52	14.82	23.86	41.96	49.37	57.60	25.51
meat	1.64	4.11	5.76	5.76	8.23	9.05	11.52	17.28	9.05
eggs	0.13	0.16	0.25	0.70	2.01	3.28	4.84	6.04	2.74
milk and milk products	5.76	10.70	23.86	49.37	83.93	145.77	167.86	236.98	84.75
butter and ghee	0.53	0.82	1.64	2.47	5.76	10.70	13.16	23.86	6.58
no. of persons per family	3.67	4.52	4.66	5.68	7.19	9.12	11.52	15.36	5.29
total no. of families	3,212	4,142	2,885	2,510	1,519	649	268	224	15,409
total no. of persons	11,788	18,712	13,443	14,247	10,919	5,916	3,088	3,441	81,554

Per capita *consumption of food items by expenditure levels.* I may give another illustration of the use of the sample survey in investigating consumption patterns. A number of enquiries were undertaken in both rural and urban areas in Bengal between 1936 and 1942. Although economic conditions and prices had changed during this period, the actual magnitude of such changes was small in comparison with that which occurred under war conditions. It has therefore been considered permissible to prepare a composite table showing the *per capita* consumption of various food items at different expenditure levels (in rupees per month per family) based on the above enquiries. Relevant data are given in Table 25. The same table has been expressed in the form of index-numbers in Table 26 in which the over-all average consumption has been adopted as 100.

TABLE 26. ESTIMATED INDEX-NUMBERS OF PER CAPITA CONSUMPTION OF VARIOUS FOOD ITEMS BY EXPENDITURE LEVEL

food items	units of measurement	monthly expenditure in rupees							
		0–10	10–20	20–30	30–50	50–100	100–200	200–300	300 and above
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
1. rice	maunds	90.8	108.9	108.4	105.5	100.2	81.7	75.4	76.9
2. rice products	,,	50.0	70.0	130.0	120.0	130.0	130.0	120.0	70.0
3. atta flour, etc.	,,	8.3	37.5	66.7	87.5	133.3	225.0	300.0	383.3
4. all cereals	,,	85.5	104.2	106.7	104.9	102.7	90.4	88.4	93.1
5. pulses	seers	47.5	85.0	112.5	110.0	100.0	95.0	105.0	110.0
6. salt	,,	64.5	86.7	91.4	91.9	102.1	102.1	112.9	118.3
7. potatoes	,,	25.4	54.2	66.1	76.3	100.0	130.5	164.4	223.7
8. vegetables	,,	34.6	28.7	61.5	69.1	101.1	146.0	170.3	217.7
9. oil	,,	38.9	66.7	77.8	83.3	105.6	133.3	144.4	172.2
19. sugar and gur	,,	13.5	37.2	50.8	59.3	98.1	223.3	148.9	172.6
11. fish	,,	25.8	32.3	45.2	58.1	93.6	164.5	193.6	225.8
12. meat	,,	18.2	45.4	63.6	63.6	90.9	100.0	127.3	190.9
13. eggs	number	4.9	5.7	13.1	25.5	73.2	119.8	177.0	220.6
14. milk and milk products	seers	6.8	12.6	28.2	58.2	99.0	172.0	198.1	279.6
15. butter and ghee	,,	8.1	12.5	25.0	37.5	87.5	162.5	200.0	362.5
index no. of persons per family	—	69.4	85.4	88.1	107.4	135.9	172.4	217.8	290.4

The consumption of all cereals is naturally fairly steady, but shows a slight falling off above the expenditure level of Rs. 100 per month. Rice products and wheat are increasingly preferred with rising expenditure level in Bengal. Index-numbers for pulses and salt are also, on the whole fairly steady. The consumption of all other items increased steadily with rising expenditure level. The disparity in consumption is very high in sugar, fish and meat, and still higher in the consumption of eggs, milk and milk products, and butter and ghee (clarified butter). Another point is worth noting—namely, the steady increase in the size of the household with rising economic status.

RADIO PROGRAMME PREFERENCE SURVEY

I shall now give an example of what is usually called "listener research" in the United Kingdom. This work was undertaken at the request of the Government of India. The object was to ascertain public reactions to the broadcast of war programmes from the All-India Radio in the earlier part of 1941. It was apprehended, however, that any direct enquiry would meet with considerable psychological resistance. In this situation the Indian Statistical Institute undertook to organize a comprehensive public preference survey with a broad coverage, including reactions to war broadcasts.

As regards the design of the survey, it may be briefly mentioned that the sample was picked out of households possessing radio licences, of which a list was supplied by the Government. The households were selected on the basis of pure space randomization, and arranged in the form of several interpenetrating replicates. Information for each replicate or sub-sample was collected by an independent set of investigators. The field work was carried out in April-May 1941, and a comprehensive report was submitted to the Government in July 1941, but it was not published owing to wartime restrictions.

Four tables are reproduced below to give a general idea of the approach. Table 27 shows the frequency (in the form of percentages) of listening "often" to different items of the radio programme broken down by age, educational, and occupational groups. The size of the sample in each group is shown at the head of each column.

The nature of the fluctuations in different age-groups is of considerable interest, and is on the whole in keeping with what may be called commonsense expectations. For example, "war news" had practically the highest preference in almost every group, which is just what might have been expected. "News talk" and "foreign news" were also, on the whole, generally popular, as these were also mostly concerned with the war.

In the music and entertainment section it was found that "modern Bengali" and "(Rabindranath) Tagore songs," as well as "plays" and "instrumental music," were, on the whole, highly popular, with "devotional music" holding quite a good place; "classical music" was, however, distinctly less popular. There were also interesting age-variations in this group. Modern Bengali and Tagore songs, plays, and instrumental music were far more popular among the younger people, and showed a definite and considerable decrease in popularity with increasing age. Devotional music, on the other hand, showed a sharp contrast and increasing popularity with age.

There are many other interesting points which, however, need not detain us. It is the sampling errors to which I should like to draw attention. The values for three independent but interpenetrating samples or replicates are shown in Table 28. The variance (calculated from three replicates—that is, on two degrees of freedom) for the different items is shown in col. 6. The "within binomial variance" for the whole material is given in col. 7; and the ratio of variances in col. 8. The population is not Gauss-Laplacian, and strictly speaking Fisher's F cannot be used to test the significance. The use of F may, however, give approximate results, and significance has been shown in the usual manner.

The important point to note is that, except in three cases, errors calculated from the replicated samples are greater than the theoretical binomial errors. This shows the need for caution.

I am giving below a table to indicate the practical use of such investigations. In this table the index-numbers of frequency of listening are shown separately for men and women in cols. 2.1 and 2.2 respectively. Index-numbers of demand (as expressed by the desire of persons interviewed to have more time given to particular items) are shown in cols. 3.1 and 3.2. Pooled index-numbers of preference were also calculated by taking the average of the two previous index-numbers, and are shown in cols. 4.1 and 4.2. Index-numbers of broadcast time actually allotted to various items at the time of the survey are shown in col. 5. Dividing the pooled index of preference in cols. 4.1 and 4.2 by the index of broadcast time we get index-numbers of disparity given in cols. 6.1 and 6.2 for men and women separately.

These index-numbers of disparity indicate to what items more time should be given. The disparity index cannot, of course, be treated as a quantitative measure of the time likely to be required for satisfaction. For example, the relative demand for talks is very high,

TABLE 27. CALCUTTA RADIO PROGRAMME PREFERENCES—APRIL–MAY, 1941 : FREQUENCY OF LISTENING "OFTEN"

item of radio programme	all groups	by age-groups (in years)					by educational groups			by occupational groups					
		14–18	19–25	26–35	36–55	55–	non-matrics	under-graduates	graduates	student	petty trade	service	business	profession	landlords
	$n = 803$	$n = 24$	$n = 170$	$n = 278$	$n = 298$	$n = 33$	$n = 156$	$n = 354$	$n = 293$	$n = 149$	$n = 11$	$n = 300$	$n = 184$	$n = 95$	$n = 64$
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
WAR															
1. war news	54.7	54.2	52.9	60.1	49.9	60.6	50.3	58.5	52.2	54.0	54.5	54.0	54.9	54.7	57.8
2. news talk	50.8	45.8	52.4	52.2	47.9	57.6	46.5	54.2	48.8	46.0	54.5	49.0	52.2	58.9	54.7
3. foreign news	43.5	41.7	44.1	46.4	39.3	54.5	35.1	46.6	45.1	46.7	27.3	41.3	41.8	49.5	43.8
MUSIC															
4. modern Bengali	46.0	70.8	59.9	47.1	37.9	18.2	45.9	49.4	42.0	58.7	54.5	41.0	44.6	40.0	51.6
5. Tagore	40.2	66.6	49.4	39.6	35.6	21.2	30.6	44.4	40.3	49.3	27.3	38.3	37.0	42.0	35.9
6. plays	42.5	54.2	51.2	37.4	41.9	36.4	51.6	51.1	33.1	48.7	45.5	42.0	45.7	38.9	53.1
7. instrumental	42.0	50.0	51.2	43.5	36.2	27.3	41.4	46.6	36.5	46.7	54.5	39.7	40.2	41.1	45.3
8. devotional	35.1	25.0	30.0	32.4	39.3	51.5	43.3	36.4	29.0	30.0	27.3	35.3	35.3	36.8	43.8
9. classical	23.3	16.7	29.9	24.8	19.8	12.1	27.4	23.2	21.2	26.6	27.3	22.0	23.4	18.9	26.6
TALKS															
10. humorous	29.8	29.2	28.8	32.4	28.2	27.3	29.3	31.6	27.6	26.7	36.4	30.7	28.3	26.3	40.6
11. scientific	26.2	37.5	27.1	28.1	22.5	30.3	20.4	24.9	30.4	28.0	9.1	27.9	21.7	27.4	26.6
12. literary	18.7	20.8	21.2	15.8	18.5	30.3	8.9	25.7	21.5	24.7	9.1	22.7	14.7	21.1	23.4
13. educational	24.9	33.3	26.5	25.2	22.5	30.3	17.8	30.2	24.6	37.3	27.3	26.3	16.8	23.2	25.0
14. music lessons	20.7	33.3	37.6	15.1	15.8	15.2	21.7	24.9	15.0	34.7	36.4	18.3	17.4	11.6	18.8
MISCELLANEOUS															
15. children's	16.1	37.5	21.8	13.3	13.4	18.2	16.6	18.9	12.3	26.0	9.1	11.7	15.8	13.7	18.8
16. women's	14.7	20.8	17.6	12.2	13.8	24.2	19.1	16.7	9.9	22.7	9.1	10.0	14.1	12.6	23.4
17. folk music	16.9	25.0	17.0	16.9	16.1	18.2	15.3	18.9	15.4	14.7	0	17.0	20.6	14.7	17.2
18. rural	12.0	12.5	18.2	11.2	8.1	21.2	10.8	12.4	11.9	16.0	9.1	11.3	12.5	7.4	10.9

but it is possible that a comparatively small increase would enable the satisfaction point to be reached. This merely means that the index of disparity has to be re-estimated as conditions change.

Finally, in Table 30 I give the reactions to war broadcasts. It is quite clear that early in 1941 enemy broadcasts were found more interesting as well as more convincing as compared to All-India Radio propaganda. The margin of error of the results (calculated from replicated samples) was reasonably small. In any case, the possibility of estimating the margin of error (by using independent networks of samples) is clearly a great gain.

TABLE 28. CALCUTTA RADIO PROGRAMME PREFERENCES—APRIL–MAY, 1941 : PERCENTAGES OF PERSONS "LISTENING OFTEN" TO DIFFERENT ITEMS OF PROGRAMME

Calcutta radio programme item early 1941	sample (replicates)			pooled values with s.e. based on replicates $n = 803$	variance		ratios of variance
	1	2	3		between samples d.f. = 2	within samples d.f. = 800	
	$n = 287$	$n = 268$	$n = 248$				
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
WAR							
war news	55.8	50.7	57.7	54.7 ± 2.52	3,340	2,479	1.35
news talk	54.0	42.9	55.6	50.8 ± 3.99	12,728	2,477	5.14†
foreign news	45.3	40.3	44.8	43.5 ± 1.59	2,032	2,461	0.83
MUSIC							
modern Bengali	48.8	39.6	49.6	46.0 ± 3.38	8,284	2,472	3.35*
Tagore	39.4	35.1	46.8	40.4 ± 3.42	8,978	2,391	3.75*
plays	43.6	41.4	42.3	42.4 ± 0.67	319	2,452	0.13
instrumental	42.2	40.7	43.1	42.0 ± 0.70	402	2,444	0.16
devotional	31.0	28.8	46.8	35.5 ± 5.67	24,734	2,225	11.12†
classical	26.1	20.1	23.4	23.2 ± 1.74	2,506	1,794	1.40
TALKS							
humorous	28.9	28.0	32.7	29.9 ± 1.44	1,567	2,094	0.75
scientific	27.9	21.3	29.4	26.2 ± 2.49	4,958	1,926	2.57
literary	24.4	16.0	14.9	18.4 ± 3.00	7,364	1,506	4.89†
educational	23.7	20.9	30.6	25.1 ± 2.88	6,450	1,861	3.47*
music lessons	18.1	19.8	24.6	20.8 ± 1.95	2,953	1,639	1.80
MISCELLANEOUS							
children's programme	14.3	14.9	19.4	16.2 ± 1.61	1,970	1,349	1.46
women's	12.9	9.3	22.6	14.9 ± 3.97	12,036	1,228	9.80†
folk music	14.3	16.0	21.0	17.1 ± 2.01	3,130	1,404	2.23
rural programme	13.2	10.1	12.5	11.9 ± 0.94	748	1,055	0.71

* Significant at 5 per cent level. † Significant at 1 per cent level.

TABLE 29. RADIO PROGRAMME PREFERENCE—CALCUTTA, APRIL–MAY, 1941 : INDEX-NUMBERS OF PREFERENCE, BROADCAST TIME AND DISPARITY BY SEXES

name of programme item	index-numbers of								
	listening		demand		preference (pooled)		broad-cast time	disparity	
	men	women	men	women	men	women		men	women
(1)	(2.1)	(2.2)	(3.1)	(3.2)	(4.1)	(4.2)	(5)	(6.1)	(6.2)
WAR	299	186	148	47	224	116	154	145	75
news talk	155	92	116	27	136	60	24	567	250
MUSIC									
modern Bengali	140	148	207	210	174	179	262	66	68
Tagore	123	128	184	236	154	182	47	328	387
plays	130	188	172	236	151	212	125	121	170
instrumental	128	114	175	176	151	144	134	113	107
devotional	107	114	142	155	124	134	219	57	61
classical	71	87	47	54	59	90	208	28	43
TALKS									
humorous	91	94	104	108	98	101	8	1,225	1,262
scientific	80	27	102	34	91	30	4	2,275	750
literary	57	47	74	27	66	37	3	2,200	1,233
educational	76	67	90	68	83	68	20	415	340
music lesson	63	101	28	94	46	98	65	71	151
MISCELLANEOUS									
children's	49	74	45	33	47	54	174	27	31
women's	45	136	38	203	42	170	73	58	233
folk music	57	40	20	7	38	24	49	78	49
rural	37	54	9	13	23	20	131	18	15

TABLE 30. WAR BROADCAST REACTIONS : CALCUTTA SAMPLE, APRIL–MAY, 1941

source of broadcast	percentage of persons finding the broadcast								
	interesting			convincing			effective propaganda		
	n	mean	s.e.	n	mean	s.e.	n	mean	s.e.
(1)	(2)	(3)		(4)	(5)		(6)	(7)	
enemy	139	36.1	2.02	133	34.5	2.18	156	40.5	2.60
allied (mostly AIR)	56	14.5	1.11	41	10.6	1.10	57	14.8	1.30
neutral (incl. USSR)	34	8.8	0.74	50	13.0	1.15	19	4.9	0.62
no opinion	156	40.5	1.74	161	41.8	2.77	153	39.7	2.72
total	385	99.9%		385	99.9%		385	99.9%	

N.B.—Standard errors based on eighteen sub-samples.

Calcutta public preference survey, 1941

I shall give another example of a different type, something like what is called public preference or "Gallup" polls in the west. The survey was carried out in certain sections of Calcutta middle-class families along with the broadcast reaction survey. The schedule covered a large number of items including preferences for classical and modern literature in both English and Bengali, different kinds of games, tobacco, *pan* (betel leaves) etc., and a number of social and political questions. A few tables are given here for purposes of illustration.

TABLE 31. CALCUTTA PUBLIC PREFERENCE SURVEY, 1941 : POSSESSION OF HOROSCOPES

groups	number of persons			difference between observed and expected	χ^2
	total	having horoscopes			
		observed	expected		
(1)	(2)	(3)	(4)	(5)	(6)
total sample	1,470	919	$p = 0.625$		
		(a) by age group			
below 18 years	38	23	23.8	− 0.8	0.07
19–25 "	259	145	161.9	−16.9	4.70
26–35 "	486	297	303.8	− 6.8	0.41
36–55 "	601	405	375.6	29.4	6.14
above 55 "	86	49	53.8	− 4.8	1.14
	$P(\chi^2) = 0.015$, d.f. $= 4$, $\chi^2 = 12.46$				
		(b) by educational groups			
non-matrics	380	195	273.5	−42.5	20.28
undergraduates	656	437	410.0	27.0	4.74
graduates	434	287	271.3	15.7	2.42
	$P(\chi^2) < 0.001$, d.f. $= 2$, $\chi^2 = 27.44$				
		(c) by occupational groups			
students	189	107	118.1	−11.1	2.78
petty trade	101	43	63.1	−20.1	17.07
service	522	329	326.3	2.7	0.06
business	390	245	243.8	1.2	0.16
learned profession	177	125	110.6	14.4	5.00
landlords	91	70	56.9	13.1	8.04
	$P(\chi^2) < 0.001$, d.f. $= 5$, $\chi^2 = 33.11$				
		(d) by expenditure levels (per month per family)			
below Rs. 40	105	45	65.6	−20.6	17.25
Rs. 41–Rs. 100	370	210	231.2	−21.2	5.18
Rs. 101–Rs. 200	386	249	241.2	7.8	0.67
Rs. 201–Rs. 400	316	216	197.5	18.5	4.62
above Rs. 400	241	173	150.6	22.4	8.88
	$P(\chi^2) < 0.001$, d.f. $= 4$, $\chi^2 = 36.60$				
		(e) by communities			
Hindus	1,389	907	868.1	38.9	4.64
Muslims	81	12	50.6	−38.6	78.98
	$P(\chi^2) < 0.001$, d.f. $= 1$, $\chi^2 = 83.62$				

Table 31 gives the percentages of persons having horoscopes, broken down by age, educational and occupational groups, expenditure levels, and communities. The total number of persons covered in the survey is shown in col. 2, and the number possessing horoscopes in col. 3. The over-all proportion of persons having horoscopes is 62.5 per cent The expected number of persons having horoscopes is shown in col. 4, and the difference between the observed and expected number in col. 5. The corresponding value of χ^2 is given in col. 6.

Variations by age-groups are not important, but non-matrics have comparatively fewer horoscopes (possibly because non-matrics on the whole belong to lower expenditure levels). The occupational distribution shows that the learned profession or landlords proportionately have more horoscopes. The distribution by expenditure levels clearly brings out that richer people on the whole have a proportionately larger number of horoscopes. Finally, from the distribution by communities it is clear that Hindus are much more interested in horoscopes than Muslims.

TABLE 32. CALCUTTA PUBLIC PREFERENCE SURVEY, 1941: REMARRIAGE OF WIDOWS AND WIDOWERS

groups	total number of persons giving opinion about remarriage of		percentage opinion in favour of							
			unconditional support		conditional support		unconditional opposition		indifferent	
	widows	widowers	widows	widowers	widows	widowers	widows	widowers	widows	widowers
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
(*a*) by educational groups (males only)										
non-matrics	367	348	19.07	22.42	52.32	58.05	21.15	13.78	7.36	5.75
undergraduates	658	634	17.02	16.56	62.31	67.67	14.44	9.78	6.23	5.99
graduates	436	410	26.15	24.15	62.16	65.60	6.19	4.88	5.50	5.37
total	1,461	1,392	20.26	20.26	59.75	64.65	13.69	9.34	6.30	5.57
(*b*) by sex (all cards)										
men	1,461	1,392	20.26	20.26	59.75	64.66	13.69	9.34	6.30	5.74
women	37	37	5.41	10.81	48.65	48.65	40.54	35.14	5.40	5.40
total	1,498	1,429	19.89	20.01	59.48	64.24	14.35	10.01	6.28	5.74
(*c*) by community (all cards)										
Hindus	1,408	1,341	17.47	17.52	60.87	65.92	15.13	10.59	6.53	5.97
Muslims	90	88	57.78	57.95	37.78	38.64	2.22	1.14	2.22	2.27
total	1,498	1,429	19.89	20.01	59.48	64.24	14.35	10.01	6.28	5.74

Table 32 shows the distribution of opinion about the re-marriage of widows and widowers by educational groups, communities, and sex. Figures for unconditional opposition to the re-marriage of widows and widowers shown respectively in cols. 8 and 9 are of considerable interest. It is clear that there is much greater opposition to the re-marriage of widows. Opposition definitely decreased with increasing educational status, and women are more opposed to re-marriages than men. Finally, unconditional opposition to re-marriage is practically restricted to Hindus.

Table 33 shows public opinion on inter-group marriages, broken down by educational groups. From col. 6 it is seen that unconditional opposition decreases in every case with increasing educational status. On the whole, there still exists considerable opposition to inter-marriage between sub-castes (28.2 per cent) which is greater against inter-marriage between different castes (36.41 per cent). It is interesting to observe that the opposition is greater (49.14 per cent) against marriages within the same *gotra* (that is, within traditionally the same patrilineal family or clan) than against marriages between provinces (44.61 per cent). Opposition against marriage between communities—*i.e.*, Hindus and Muslims (60.64 per cent)—is as strong as opposition to marriages between different nationalities (58.66 per cent). It is interesting to note, however, that on the whole 12 or 13 per cent of the persons surveyed and about 20 or 22 per cent of graduates are unconditionally in favour of inter-communal or international marriages.

TABLE 33. CALCUTTA PUBLIC PREFERENCE SURVEY: INTER-GROUP MARRIAGE BY EDUCATIONAL GROUPS

type of inter-group marriage	educational groups	total number	percentage of			
			unconditional support	conditional support	unconditional opposition	indifferent
(1)	(2)	(3)	(4)	(5)	(6)	(7)
(*a*) between sub-castes	non-matrics	367	19.35	27.79	38.42	14.44
	undergraduates	658	31.00	27.66	30.55	10.79
	graduates	436	44.50	30.96	10.06	8.48
	total	1,461	32.10	28.68	28.20	11.02
(*b*) between castes	non-matrics	367	17.17	22.88	45.78	14.17
	undergraduates	658	23.56	25.84	40.42	10.18
	graduates	436	38.53	30.05	22.48	8.94
	total	1,461	26.42	26.35	36.41	10.82
(*c*) within same *gotra*	non-matrics	348	12.64	15.23	54.89	17.24
	undergraduates	635	16.70	20.94	51.18	11.18
	graduates	409	26.40	21.03	41.08	11.49
	total	1,392	18.53	19.54	49.14	12.79
(*d*) between provinces	non-matrics	348	10.92	22.99	53.16	12.93
	undergraduates	635	15.91	27.87	47.87	8.35
	graduates	409	26.65	30.32	32.27	10.76
	total	1,392	17.82	27.37	44.61	10.20
(*e*) between communities (Hindus and Muslims)	non-matrics	367	8.17	10.90	67.03	13.90
	undergraduates	658	9.57	14.13	66.42	9.88
	graduates	436	22.25	19.50	46.55	11.70
	total	1,461	13.00	14.92	60.64	11.44
(*f*) between nationalities	non-matrics	367	6.27	9.53	67.85	16.35
	undergraduates	658	10.03	17.17	62.16	10.64
	graduates	436	19.50	21.79	45.64	13.07
	total	1,461	11.91	16.63	58.66	12.80

Finally, Table 34 shows public opinion about religious instructions in colleges, broken down by educational groups. From col. 6 it is seen that 59.34 per cent are in favour of such instruction and 18.83 per cent not in favour, while nearly the same proportion appears to be indifferent. From col. 7 it is clear that opinion in favour of religious instruction in colleges definitely decreases with increasing educational status and that nearly one-third of graduates are not in favour of such instruction.

TABLE 34. CALCUTTA PUBLIC PREFERENCE SURVEY, 1941: RELIGIOUS INSTRUCTION IN COLLEGES BY EDUCATIONAL GROUPS

educational group	number of persons				percentage of persons			
	in favour	not in favour	indif-ferent	total	in favour	not in favour	indif-ferent	total
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
non-matrics	288	22	87	397	72.54	5.54	21.92	100.00
undergraduates	422	116	125	663	63.65	17.50	18.85	100.00
graduates	179	144	115	438	40.87	32.88	26.25	100.00
total	889	282	327	1,498	59.34	18.83	21.83	100.00

MODEL SAMPLING EXPERIMENTS

I must refer to another special feature of the Institute—namely, the extensive use of model sampling experiments in applied work. For example, in the Bengal crop survey, information was collected about the crop grown on each individual field for large tracts, and these were recorded on maps. It was then possible to carry out, in the Statistical Laboratory, a large variety of model sampling experiments using grids of different sizes and densities. Such studies proved most helpful in developing the actual sampling technique.

In the case of space distributions, it has been shown elsewhere that the patch number supplies a useful concept. Consider a map of, say, an agricultural area with fields which are assumed, for the sake of simplicity, to be of the same size. Suppose each field which is shown with paddy is coloured yellow, and fields without paddy are left blank or white. The map of any such region would be then broken up into a number of yellow and white patches, depending on the actual distribution over space of fields under paddy. If alternate fields in both directions are sown with paddy, then the pattern will be something like that of a chessboard; and in this case the total number of patches will be equal to the total number of fields. On the other hand, if half the whole area is sown with paddy in one compact block and the other half is left without paddy, then the patch number will be simply two. In any actual situation the patch number will lie between these two limits. If the fields to be sown with paddy (or to be marked yellow) are selected purely at random, then the patch number would show a characteristic distribution.

It is obvious that the design of the sample survey can be made more efficient if something is known about the distribution of the patches. For example, if the patch-pattern is of the chessboard type, then it will be quite sufficient to explore (in the ideal limit) only two adjoining cells to secure a complete picture. If the patch number is only two, then the best plan would be to try to settle the boundary between the two heaps. In any case, when there is any appreciable tendency towards patterning, it is often possible to have recourse to configurational (which is sometimes called "systematic") sampling.

For two colours (*i.e.*, for a binomial distribution) in one dimension the theoretical solution is known. For more than two colours in one dimension, or for two or more colours in two dimensions, complete theoretical solutions have not yet been given. Raj Chandra Bose is carrying on mathematical researches on the sampling distribution of the patch number in the case of a binomial field in two dimensions, and has obtained a number of useful results.

Investigations are also proceeding on purely experimental lines under the leadership of Jitendra Mohan Sen Gupta. A number of different sets of fields each of size 100 × 100 —*i.e.*, consisting of 10,000 cells arranged in the form of a square—were constructed and filled with numbers from 0 to 9 at random (or, to be more strict, what are believed to be arranged at random) in two dimensions. Some were based on well-known tables of random numbers. A large number were produced mechanically with the Hollerith equipment by the British Tabulating Machine Co., Ltd. By assigning different digits to different colours, it is possible to prepare random distributions of patches on these fields and to study the patch number experimentally. Work is being done on a large scale, and has already yielded results which, it is believed, would supply useful information for the guidance of practical work, as well as valuable clues for mathematical researches.

I shall give one concrete example. Consider an $n \times n$ field. Let p be the probability for a cell to be black, and q to be white $(p + q = 1)$. In counting black patches contact is recognized only along sides, but in counting white patches contact is permitted both along sides and at corners. Let B be the number of black patches and W the number of white patches (as defined above), but embedded within black patches. R. C. Bose has given the following results :

$$\text{Expectation } E(B - W) = p + 2(n-1)pq + (n-1)^2(pq^2 - p^3q)$$

$$\begin{aligned}\text{Variance } V(B - W) = {} & n^2pq - 4(n-1)p^2q - (6n^2 - 10n + 4)p^2q^2 - (6n^2 - 40n + 50)p^3q^2 + \\ & + (14n^2 - 56n + 54)p^3q^3 + (32n^2 - 104n + 84)p^4q^3 - (9n^2 - 30n + 25)p^4q^4.\end{aligned}$$

The results of one series of model sampling experiments are given in the following two Tables, 35 and 36. The observed average number of (black minus embedded white) patches and the expected number calculated by the above formula are shown in cols. 2 and 3 of Table 35 for various values of p. The difference between the observed and expected number divided by the standard error of the difference (based on the theoretical value of the variance as calculated from the formula given above) is shown in col. 4. Although the distribution may not be strictly normal in the present case, it is likely that the figures in col. 4 may behave approximately as normal deviates. It will be noticed that out of thirty-six comparisons no fewer than ten are significant at 1 per cent level and three at the 5 per cent level.

TABLE 35. MODEL SAMPLING EXPERIMENTS: COMPARISON OF EXPECTED AND OBSERVED AVERAGE NUMBER OF (BLACK MINUS EMBEDDED WHITE) PATCHES IN A BINOMIAL FIELD

p	average number observed	average number expected	normal deviate	average number observed	average number expected	normal deviate
(1)	(2)	(3)	(4)	(5)	(6)	(7)
	10 cells × 10 cells (n = 1000)			20 cells × 20 cells (n = 250)		
0.1	7.99	8.21	−3.07†	31.54	32.44	−0.35
0.2	12.81	12.93	−1.56	49.78	50.18	−1.34
0.3	14.44	14.46	−0.32	54.83	54.52	0.95
0.4	13.30	13.21	0.32	48.34	47.67	1.76
0.5	10.28	10.06	2.17*	33.62	32.56	2.51*
0.6	5.29	5.70	−4.09†	11.97	13.19	−2.93†
0.7	0.96	1.25	−3.38†	− 7.37	− 5.72	−4.70†
0.8	−2.06	−2.02	−0.63	−18.34	−18.53	0.74
0.9	−2.89	−2.66	−4.85†	−18.34	−18.75	2.03*
	50 cells × 50 cells (n = 200)			100 cells × 100 cells (n = 50)		
0.1	199.93	201.24	−1.67	797.48	802.98	−1.76
0.2	305.87	307.84	−2.40*	1,215.70	1,233.68	−4.46†
0.3	328.44	328.45	—	1,294.34	1,297.93	0.98
0.4	278.42	277.47	0.87	1,090.06	1,082.91	1.64
0.5	177.40	175.06	1.93	681.74	662.56	3.93†
0.6	41.03	47.17	−5.17†	122.82	142.21	−4.05†
0.7	− 77.50	− 74.52	−3.00†	−352.26	−348.78	0.87
0.8	−151.82	−152.55	1.03	−654.86	−657.71	1.01
0.9	−143.17	−143.70	0.92	605.84	−607.50	0.71

* Significant at 5 per cent level. † Significant at 1 per cent level.

A similar comparison of observed and expected values of the standard deviation is shown in Table 36. Using large sample theory, the differences between observed and expected values are found to be significant in three out of thirty-six cases at the 5 per cent, and three out of thirty-six cases at the 1 per cent level of significance.

TABLE 36. COMPARISON OF OBSERVED AND EXPECTED VALUES OF STANDARD DEVIATIONS OF THE NUMBER OF (BLACK MINUS EMBEDDED WHITE) PATCHES IN A BINOMIAL FIELD

p	10 cells × 10 cells (n = 1,000)			20 cells × 20 cells (n = 250)			50 cells × 50 cells (n = 200)			100 cells × 100 cells (n = 50)		
	s.d.		normal deviate	s.d.		normal deviate	s.d.		normal deviate	s.d.		normal deviate
	observed	expected		observed	expected		observed	expected		observed	expected	
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
0.1	2.24	2.27	−0.60	4.31	4.47	−0.80	11.17	11.08	0.16	21.85	22.09	−0.11
0.2	2.42	2.40	0.40	5.20	4.70	2.38*	12.90	11.62	2.19*	25.71	23.18	1.09
0.3	2.62	2.55	1.17	5.75	5.13	2.70	15.20	12.89	3.61†	30.26	25.84	1.71
0.4	2.99	2.91	1.14	6.05	6.01	0.15	15.67	15.34	0.43	32.16	30.89	0.41
0.5	3.16	3.20	−0.57	6.35	6.68	−1.10	17.45	17.12	0.38	41.72	34.52	2.09*
0.6	3.19	3.17	0.29	6.38	6.58	−0.69	16.76	16.81	−0.06	29.71	33.85	−1.22
0.7	2.66	2.72	−1.00	5.55	5.55	0.00	13.57	14.03	−0.66	25.16	28.16	−1.06
0.8	1.82	1.99	−4.25†	3.92	4.01	−0.50	10.05	10.00	0.10	17.85	19.98	−1.06
0.9	1.70	1.50	6.67†	3.36	3.19	1.21	8.49	8.23	0.63	17.01	16.64	0.22

* Significant at 5 per cent level. † Significant at 1 per cent level.

It is clear that expected values of the patch number or of the variance are not confirmed in a rigorous manner. The general agreement between observed and expected values indicates, however, that the theoretical results are probably not wrong. The failure is, therefore, most probably due to some or all the fields used for experimental sampling not being of a sufficiently random character. This often happens in practice. In fact, as pointed out in the paper on large-scale sample surveys (in the *Phil. Trans.*), the concept of degrees of randomness is particularly useful in situations similar to the present one. I may conclude this section by stating that we have a big programme of work for model sampling experiments, but progress is slow for lack of resources.

SAMPLE SURVEY OF THE ECONOMIC BACKGROUND OF THE BENGAL FAMINE

A sample survey was undertaken in 1944-45 to collect information relating to the after-effects and the economic background of the Bengal famine of 1943. A first report by Ramkrishna Mukherjee, Ambika Ghosh and myself has been published in *Sankhyā* (the Indian Journal of Statistics), Vol. 7, Part 4, and it is not necessary to enter into details, but a few typical results may be of interest in the present connection. The enquiry covered 15,769 families selected at random from 386 villages, which themselves were selected at random from 41 (out of 86 rural) sub-divisions (administrative units) covering about 60,000 sq. miles in Bengal. The design was zonal, with stratification of sub-divisions by intensity of incidence of famine conditions; and randomization of villages was completed separately within each sub-division.

From the sample survey it appeared that the land position was precarious even before the famine, with one-third of all rural families having no paddy land, while two-fifths had less than 2 acres, so that about three-fourths of all rural families had either no paddy land or owned less than 2 acres. With an average production of 820 lbs. of rice per acre, an average consumption of about 320 lbs. per head per year and an average family size of 5.4 persons, the subsistence level would be about 2 acres of paddy land per family on an average. The actual over-all average for the province was, however, found by the sample survey to be about 1.8 acres of paddy land per rural family, which was below the subsistence level. It is not surprising therefore that, averaged over a number of years before the war, there was a small but net import of a little over 1 per cent of total production of rice and other cereals into the province.

The sample survey showed that the classification of sub-divisions by amount of paddy land owned per family before the famine was roughly parallel to the degree of incidence of famine conditions, indicating that sub-divisions in which there were more families with paddy land below subsistence level were more vulnerable to the famine.

It was found that about 1.6 millions of families (about one-fourth of the number who had owned paddy land before the famine) had either sold in full or in part, or had mortgaged their paddy land during the famine period. About a quarter of a million of families were obliged to sell all their paddy land, and were thus reduced to the rank of landless labour.

The net loss of plough-cattle was about a million, or 13 per cent, during the famine period. Sales of cattle largely exceeded purchases, showing that transfers had taken place not merely from one rural family to another, but that large purchases to the extent of about 600,000 head of cattle had been made by outsiders (possibly by contractors for the supply of meat for army consumption).

The economic deterioration was measured by the number of families transferred from occupations at a higher economic level to occupations at a lower level. Assessed by such methods, it was found that about 700,000 of rural families had suffered a lowering of economic status during the famine. It was also found that even at the beginning of 1943 (before the advent of famine conditions) there had been already an increase of 150,000 of destitute persons. There was a further increase of about 330,000 destitutes during the famine itself, so that nearly half a million persons were rendered destitute under war and famine conditions in Bengal.

Even in the pre-famine period (January 1939 to January 1943) about 6.84 per cent of rural families had suffered economic deterioration against 3.32 per cent who had improved their economic level, while the position of 1.07 per cent was not clear. Economic deterioration had thus set in definitely in the pre-famine period. Rates of change became more rapid during the famine period. Improvement in economic status during the famine period was relatively twice as great as that in the pre-famine period; but this was offset by a three times greater rate of economic deterioration and twelve times greater of destitution. The famine period was thus one of greatly accelerated economic changes. Improvement of economic conditions was quicker, but was restricted to a comparatively small number of families. Deterioration and destitution were even more accelerated, and were shared by a large number of families. The poorer sections of the community, especially landless labour, fishermen, and village craftsmen, were most seriously affected, and many were rendered destitute; the middle group, who had land of their own and other assets, were naturally less vulnerable; and a comparatively small number of families in the upper stratum had remained immune, and sometimes even became prosperous.

The above summary gives an idea of the kind of information which it was possible to secure by the sample survey. The margin of error of the sample estimates in the cases investigated varied roughly between 3 and 5 per cent. Comparison with other available statistics also showed differences of the same order. For example, extrapolating from the cattle census figures of 1930 and 1940, the calculated number of plough-cattle in rural Bengal in the beginning of 1943 was about 8.3 millions. The sample estimate was about 7.9 millions, giving a difference of about 5 percent. The sample estimate of 18.6 millions of acres of paddy land was also not inconsistent with other available estimates. On the whole, the above enquiry showed the possibility of using the sample survey in a quick and efficient manner to obtain information about economic conditions in rural areas with a margin of error sufficient for many practical purposes.

A MULTI-PURPOSE SURVEY IN BENGAL, 1946–47

I may briefly refer to an extensive survey (which is actually in progress in Bengal at the present time, in 1946–47) as an example of a multiple survey. The size of the grid (sample unit) in this case is just 1 sq. mile, and the information is being collected in the form of two independent but interpenetrating networks of sample units. In the first stage of the survey all households falling within each grid were surveyed with a rather short schedule covering a number of basic items, such as sex and age composition of the family, caste and community, occupation, agricultural land, number of cattle, total indebtedness, etc. Information has been already collected for 84,370 families from 475 grids scattered over about 60,000 sq. miles of rural Bengal. The information is being tabulated on Hollerith machines, and certain preliminary estimates have been made on the basis of about half the number of families surveyed.

The total rural population based on this portion of the material is found to be 50.16 millions, roughly as in March or April 1946, with a calculated standard error of 5.26 millions. The corresponding census population in rural Bengal in March 1941 was 47.185 millions. Owing to uncertainties in census counts it is difficult to calculate reliable rates of growth of population, especially in Bengal. It is also difficult to assess the exact effects of the Bengal famine of 1943 on the growth of population. We may, however, make some rough calculations. The total population (including rural and urban areas) of Bengal was about 46.7 millions in 1921, 50.1 millions in 1931, and 60.3 millions in 1941. Very rough estimates by linear extrapolation can be made on the above basis. Adopting the rate of growth during 1931–41, the extrapolated rural population as in March 1946 would be about 51.9 millions. There are, however, reasons to believe that the census count in 1931 was abnormally low on account of the non-co-operation movement. Adopting the average rate of growth during the twenty-year period 1921-41, the extrapolated figure for the rural population at the end of 1944 would be about 50.4 millions. The observed sample estimate of 50.1 millions is thus of the right dimensional order; and is fairly satisfactory with a sampling fraction of about 1 : 250.

The enquiry just completed is, however, only the first fold or layer of the survey. Arrangements are now being made to select about 20 per cent of the families included in the first survey for a more detailed study of rural indebtedness and agricultural labour. If everything proceeds smoothly it should be possible to amplify the results for the second survey for making estimates for the sample covered in the first survey, and hence for the rural population of Bengal as a whole. After the second survey is over, if funds permit, it is proposed to reduce the number of families much further and carry out a detailed enquiry into family budget and other socio-economic conditions. At each successive survey the families would be selected at random, but with appropriate zoning and/or stratification, so that it should be possible to amplify the results of one survey to obtain estimates for the families covered in an earlier survey, and hence, finally, for all rural families in the province.

OTHER WORK IN STATISTICAL SAMPLING

A good deal of other work on statistical sampling has been and is being done in the Indian Statistical Institute. In most cases these have definite practical ends in view. The important point which I should, however, like to emphasize is that such practical studies have continually given rise to theoretical problems of fundamental interest. The subject of crop surveys itself raised many questions of theoretical interest, to some of which I have already referred. I have also mentioned that probability problems relating to space distributions are being tackled on fruitful lines with the help of topological and combinatorial concepts by Raj Chandra Bose. The close integration of applied work and theoretical researches has been such a valuable feature of the work of the Institute that it seems worth while giving a few more examples.

Analysis of anthropometric measurements. The statistical analysis of anthropometric measurements—work which may be properly considered to be purely of an applied nature—had led, about 20 years ago, to interesting theoretical developments in the formulation of the generalized distance (D^2-statistic). It was the starting point of a good deal of mathematical research in the theory of sampling from multivariate correlated population by Raj Chandra Bose and Samarendra Nath Roy, whose work in this subject is already well known.

RECENT EXPERIMENTS IN STATISTICAL SAMPLING IN I.S.I.

We had occasion recently to take up an anthropometric study of an extensive series of physical measurements covering 3,000 individuals belonging to twenty-two different castes and tribes in the United Provinces of India. All the measurements were taken by one individual observer, Dr. D. N. Majumdar of the Lucknow University, and were therefore free from differences of personal equation. This made the material particularly suitable for comparative purposes. The use of the D^2-statistic has, I believe, yielded information of great interest and significance. As the paper is in the press, and will be shortly available, I need not enter into details.

I may mention one result which is of some methodological interest. From general considerations I reached the conclusion that, when correlations between variates are taken properly into consideration (as in the D^2-statistic), the use of indices (such as the cephalic or the nasal index) do not supply any additional information. This has been fully confirmed by C. R. Rao by actual numerical calculations. In fact, the present study has given rise to a number of interesting theoretical developments on which Rao has done some very useful work.

More recently, in 1945–46, Dr. D. N. Majumdar, working under the auspices of the Institute, has collected a large volume of material relating to physical measurements and blood-group tests of about 4,000 individuals belonging to various communities, castes, and tribes of Bengal. The analysis of this material should yield valuable results. A survey of blood pressure which was concluded some time ago is perhaps worth mentioning in the present connection.

Design of experiments. At one time a great deal of work relating to the design of agricultural experiments has been done in the Institute. This had led to significant theoretical researches in which the concepts of the Galois field and finite geometry were used with great success by R. C. Bose and, under his leadership, by workers like K. Raghavan Nair, K. Kishen, C. R. Rao, Harikinkar Nandi and others.

Circulation of rupees and rupee notes. At the request of the Reserve Bank of India, experimental surveys were made and methods were devised for estimating the circulation of silver rupees of various dates by sample counts at a number of receiving centres, like banks or railway stations. A method was also devised, based on sample counts of rupee notes received back in the Reserve Bank, to estimate the average life of such notes. It may be mentioned, incidentally, that this project gave rise to an interesting problem of measuring the magnitude of the difference between samples drawn from multinomial populations which was tackled on fundamental lines on the theoretical side by Anil Kumar Bhattacharyya, with promising results. A brief note was issued in the Report of the Reserve Bank of India on Currency and Finance for the year 1940–41, pp. 49–54.

Sampling for yield of cinchona bark. I have already mentioned the project for estimating the yield of cinchona bark as an example of multiple sampling. It is interesting to note that the final choice of physical measurements was made in this case from considerations of cost. It was found, for example, that although the measurements of the surface area of the standing plant gave the highest correlation with the yield of bark (0.865 in one case against the next highest 0.690 for a different character), the use of three simple physical measurements like the "standing vertical height of the plant," the "girth at height of 6 inches above ground level," and the "number of stems of the plant at ground level," gave a multiple correlation of 0.848. Theoretically the use of surface area would be more efficient, inasmuch as observations on eighty-nine plants would give the same information as the measurement of the three other characters on 100 plants. The time and trouble to measure the surface area for eighty-nine plants would be, however, far greater than that for measuring the other three characters on 100 plants. This is a simple example of the use of cost considerations in settling the sampling programme.

This particular project also has given rise to many interesting problems of the estimation of appropriate errors in the case of multiple sampling of various kinds. Mrs. Chameli Bose has obtained a number of useful results and is working on the subject.

Population enquiries. After the Indian Census of 1941, the Government of India had decided, as a matter of economy, to cut down most of the standard tables; even age or occupational tables were not prepared. Fortunately, Mr. M. W. M. Yeatts, as Census Commissioner, had issued instructions for the preservation of a 2 per cent sample consisting of every fiftieth individual census slip. In certain areas, mostly in Indian States, full tabulation had been carried out. A comparison of results based on the complete count and on the 2 per cent Y-sample in such areas showed that it should be possible to reconstruct most of the tables with sufficient accuracy for practical purposes. About 7 millions of the 2 per cent slips have been brought over to the Statistical Laboratory in Calcutta from all over India, and the work of transferring the information to Hollerith cards has already started. In only one province—namely, Bihar—the original census slips had also been preserved; and *ad hoc* sampling studies on the basis of this material are also proceeding.

The above projects were sanctioned on the recommendations of the Population Data Committee which was appointed by the Government of India in May 1944, and which submitted its report in June 1945. A comprehensive programme relating to population census and demographic statistics generally was prepared by this Committee, and it was definitely recommended that a continuing sample census of the Indian population should be carried out from year to year.

Statistics of road development. Mention may also be made of the use of sampling methods for traffic count on roads and connected socio-economic variates in the traffic catchment. The object in this project is to lay down a scientific foundation for preparing programmes of road development. K. B. Madhava and Satyabrata Sen are at present actively engaged in this work.

Postal traffic and revenue. An enquiry is proceeding at present to explore the possibilities of using the sampling method for estimating in advance the volume of traffic and revenue in different postal sectors. Exploratory surveys under the guidance of Satyabrata Sen have already yielded encouraging results.

INTEGRATED PROGRAMME OF WORK

The need of handling computational work on a very large scale has been salutary in teaching us the importance of efficient organization of the human agency. J. M. Sen Gupta and N. T. Mathew and others have done valuable work in this direction. An account of this work, especially from the point of view of cost accounting, would, I think, be of considerable interest. Unfortunately, we have not yet had sufficient time or money to undertake this in a systematic manner.

Along with computational work may be mentioned the preparation of statistical tables of various kinds (which often require the calculation of numerical values of Bessel and other mathematical functions), which has been proceeding for many years, and offers scope for training in a very useful branch of statistical work. The names of J. M. Sen Gupta, Purnendu Kumar Bose, and Raja Rao may be mentioned.

The contact between statisticians and field workers has also been fruitful in every way. The credit for the successful organization of the field side is almost entirely due to the organizing leadership and the ability to appreciate statistical needs on the part of Nihar Chandra Chakravarti and workers in the Field Branch, like Dhirendra Mohan Ganguly, Pronay Kumar Chatterjee and others. Much valuable work on the field has also been done by statisticians like J. M. Sen Gupta, Birendranath Ghosh, N. T. Mathew, Purnendu Kumar Bose, Harikinkar Nandi, and others. I may perhaps also mention that during the last few years the usual practice was for all theoretical workers, including R. C. Bose and S. N. Roy, to live in camp on the field for a few weeks in connection with the crop-survey work.

Training in statistics has also been an important part of Institute activities from the very beginning. One-year post-M.Sc. courses, as also occasional special courses in particular subjects, are being arranged for a long time. The Institute workers have also been intimately associated with the two-year course leading to the M.A. or M.Sc. degree and the recently introduced two-year course leading to the Honours B.Sc. degree in Statistics in the Calcutta University.

In the light of the experience gained during the last fifteen years I am convinced that the close integration of applied work, theoretical research, and training offers the soundest line of advance in statistical work, especially on the professional side. It offers scope for developing the spirit of team work among persons with widely varying interest which alone can make it possible to tackle many problems on a scale large enough to ensure success. I have been fortunate in having secured a large group of able and enthusiastic workers, each of whom in his own line has much greater knowledge and ability than myself. Whatever success the Indian Statistical Institute has achieved is in fact essentially due to the spirit of co-operative effort.

REFERENCES

BHATTACHARYYA, A. (1946) : On a measure of divergence between two multinomial populations. *Sankhyā*, **7**, 401–6.

DEMING, W. E. (1944) : Errors in Surveys. *American Sociological Review*, **9**, 359–69.

HUBBACK, J. A. (1927) : Sampling for rice yield in Bihar and Orissa. Imperial Agricultural Research Institute, Pusa, Bulletin No. 166, and reprinted in *Sankhyā*, **7**, 281–94.

KING, A. J., AND JESSEN, F. R. (1945) : The master sample of agriculture. *J. Amer. Stat. Ass.*, **40**, No. 229, 38–56.

MAHALANOBIS, P. C. (1941) : A statistical report on the rupee census. Published in the report on currency and finance, 1940–41, by Reserve Bank of India, 49–55.

——— (1942) : Sample surveys. Presidential address, Section of Mathematics and Statistics. *Procedings*, Indian Science Congress.

——— : On large scale sample surveys. *Phil. Trans.*, **231** (B), No. 584, 329–451.

———, MUKHERJEE, R. K., AND GHOSH, A. (1946) : A sample survey of after-effects of Bengal famine of 1943. *Sankhyā*, **7**, 337–400.

SUKHATME, P. V. : Bias in the use of small size plots in sample surveys for yield. *Nature*, **157**, No. 3993, 630.

WALD, A. (1945) : Sequential tests of statistical hypotheses. *Ann. Math. Stat.*, **16**, 117–86.

YATES, F. (1934) : Some examples of biassed sampling. *Ann. Eugen.*, **6**, 202–13.

——— (1945) : A review of recent statistical developments in sampling and sampling surveys. (Read before the Royal Statistical Society, January 23rd, 1945.)

Discussion on Professor Mahalanobis's Paper

Dr. F. Yates : I have much pleasure in proposing the vote of thanks. Professor Mahalanobis has given us a most interesting description of the work for which he has been responsible in India. His address has provided a vivid picture of the Institute which he directs, and at the same time has shown what enormous strides have been made there, and how extremely able his direction has been.

There is not time to discuss all the many interesting points that Professor Mahalanobis has raised, but I should first like to take the opportunity of expressing my admiration for the way in which he has handled the analysis of the social survey and broadcast survey data, the discussion of which formed the last part of his address. Those who have had experience of the analysis of data of this type will know that while it is comparatively easy to present tables giving over-all percentages, etc., the critical analysis of such material does present very great difficulties. Professor Mahalanobis is certainly to be congratulated on the penetrating methods of analysis he has developed in connection with these surveys. There is no doubt that the statistical analysis is the most difficult part of social survey work of this kind, and it is much to be hoped that further efforts will be made to develop the methodology.

Professor Mahalanobis did not spend much time discussing his sampling methods. There are several points of mutual interest that I should like to have the opportunity of discussing with him personally, but which I will not trouble to bring up now. I fully agree with Professor Mahalanobis that the determination of the sampling procedure is often the easiest part of a survey, but I would stress that this ease is in fact based on experience, which has to be acquired by a thorough study of the variability of the material that is being sampled.

Professor Mahalanobis finds the sampling problems easy because his Institute has built up a body of experience in sampling of the particular types of material with which it is concerned. This contains a moral for us all. We cannot devise good statistical methods merely by sitting in our studies and theorizing. Good statistical methods almost invariably result from contact between the mathematical statistician and the workers who are responsible for collecting the data, and who are interested not in the statistical methods themselves, but in the conclusions that emerge from their work. Consequently, the statistician must be in intimate contact with the actual numerical material. He must not think solely in terms of algebraic symbols. Professor Fisher once made what I consider a most revealing remark : "Most of the statistics which I have learnt, I have learnt on the computing machine."

Professor Mahalanobis has been concerned with the problem of securing numerical accuracy in the computations. In this connection I might mention the experience I had in the geodetic survey of West Africa. In that part of the world it was necessary to carry out much of the main framework of the survey by means of traverses instead of triangulation. The difficulty with traverses is that they involve a vast amount of computation which is not self-checking. In triangulation, once the base line is determined, the remainder of the computations are self-checking—it is impossible to make an error of any magnitude in a computation which will not reveal itself later in glaring discrepancies.

We had a very great struggle to get the numerical work of our traverses correct. We eventually evolved a most rigorous system of computation. All computations were done in the field in duplicate, the duplicates being compared only at certain points. A third computation was then done at headquarters using the "setting-out" taping, which was carried out in metres instead of feet, and bearings 45° different from the true bearings, so as to give different sines and cosines. This last computation was carried out so as to eliminate errors of the type that are likely to be made independently by different computers—*e.g.*, writing an odd taping length of 449.321 ft. as 499.321 ft., when practically all the taping lengths are just under 500 ft.

I was interested, after this system had been evolved, in finding that the U.S. Coast and Geodetic Survey laid it down in their standard instructions that triangulation was to be used in preference to traverses, if the cost was not more than 100 per cent greater, simply because of the danger of computing errors. (It is now recognized that with modern invar tapes traverses are quite as accurate, possbily more accurate, than triangulation.)

Professor Mahalanobis stressed the importance of providing the results of sampling surveys on the required dates. We had an interesting example of this—which illustrates the flexibility of sampling methods—during the war, when we were making an estimate of the amount of standing timber in the country. The first estimates were required within about five months of the decision to undertake the survey. Initially the survey had been planned to be undertaken in two 5 per cent samples, the sampling units being 6-inch Ordnance Survey Quarter Sheets (3 miles × 2 miles). Owing to the almost inevitable delays in organization, collection of staff, etc., it was clear by the time we were ready to start work, that the first 5 per cent sample would not be completed and analysed by the given date. I therefore insisted that the first 5 per cent sample should be divided into two parts. This was done by the simple process of drawing a diagonal line across each sampled 6-inch Ordnance Survey Quarter Sheet. There was, in fact, considerable opposition to this, as it was thought that the results of a 2½ per cent sample would not be sufficiently accurate for the required purpose. Fortunately I knew, from previous experience, the variability of the material, and was able to promise that the results would be of the required accuracy, and that moreover when they were obtained, estimates of the actual standard errors could be furnished so as to provide concrete evidence of this accuracy. I am not sure whether these standard errors really carried conviction, but what certainly did convince those who were initially in doubt was the close agreement between the first 2½ per cent sample and the second 2½ per cent sample, and between these and the second 5 per cent sample.

This illustrates an important point in survey work. In general, over-all estimates for the whole area are first required. It is on these estimates that the broad administrative decisions are based. Subsequently more detailed estimates for different parts of the area may well be required, in order to implement the administrative decisions. By carrying out a sampling survey in stages, as was adopted in the Forestry Survey, and as is also, I think, Professor Mahalanobis's practice, the administrators can be given the necessary information on which to base their decisions, while additional data for more detailed and accurate estimates are being collected.

Professor Mahalanobis mentioned the large biases that he has encountered when taking small areas for estimating crop yields. To a certain extent, we have had the same trouble in this country, but only to a limited degree. I suspect that many of these troubles will gradually fade away, as a permanent organization of trained people is built up. In our survey work in Africa, for example, much of the field work was undertaken by native surveyors who were trained in a Survey School which had been specially set up for the purpose. We had very little trouble with these surveyors. The very rigorous methods of booking observations which are always followed in survey practice were a great help. All observations must be booked direct into properly printed field note-books which have numbered pages. The surveyor is completely at liberty to reject any observations with which he is not satisfied, but a permanent record of such observations is always available, so that it is possible to find out what the surveyor has really done. Sometimes they did very funny things. One surveyor—quite a conscientious man—had been taken to task because his field work was not reaching the required standard of accuracy. Subsequent to this he turned in a series of traverses in which the closing errors in his bearings (on astronomical azimuths) were far smaller than they should have been. At first it was thought that the results had been deliberately faked. Subsequent cross-questioning revealed, however, that having made a preliminary computation of his closing error and finding it larger than he thought it should be, he had gone back the next day and re-observed some of the angles, rejecting the old or the new observation of each angle according to which improved the closure. The methods of booking enabled him to prove that this was what had been done, and also enabled the matter to be set right without any re-observation.

DR. C. OSWALD GEORGE, in seconding the vote of thanks : Never have I so much enjoyed listening to an address of this character, not only because of its intrinsic interest, but also because of the particularly charming manner of its delivery.

The first point which strikes one is perhaps the difference between conditions in India and in this country. On the other hand, there are important similarities. Of the various headings Professor Mahalanobis has written on the blackboard, we in this country also have to consider problems of time and sampling errors, and —for there is a body here known as the Treasury—the cost of statistical enquiries. Calculating and recording errors are also not unknown here. Had I been speaking a fortnight ago, I might have spoken more confidently on the much higher standards of accuracy in this country, but I fear Professor Mahalanobis may have seen the unfortunate calculations of mean yields in a recent issue of what we had previously regarded as an almost infallible newspaper. So I will pass over that point.

What he calls errors from physical fluctuation, I assume, arise with crops such as jute, where one reporter may find two acres under the crop, while another reporter, visiting the same farm a fortnight later, may find no jute acreage whatever. Under the existing system in this country, where returns relate to a specified day, the same problem does not arise, but some of his sampling problems interest us. He tells of the various sampling methods he adopted, "unitary and zonal, configurational and unrestricted," but unfortunately he has not given us sufficient detail to permit any useful comment on their theoretical or practical merits. He frequently mentioned large-scale sampling theory, but if greater use had been made of stratified sampling, small sample theory would possibly have come more into the picture.

I was particuarly interested by his detailed treatment of errors, for I think the question of errors (other than sampling errors) is of more importance than is sometimes believed. How many can lay their hands on their hearts and say that in their published work, errors—if not quite of the type most troublesome in India, nevertheless of serious import—have never reached the printers? But it is noteworthy that Professor Mahalanobis does not mention errors due to the machines : perhaps in the Indian climate machines are infallible!

Another point that struck me was the speaker's attitude towards a system of complete enumeration and his declaration, when comparing it, detrimentally, with sample surveys, that complete enumeration is entirely lacking in any kind of control and precludes any valid estimate of the margin of error of the final results. Without further explanation it is not easy to see why the underlying weaknesses are inherent in the one and entirely absent from the other, or why a system of duplicated surveys cannot be used in conjunction with complete enumeration. And whatever may be desirable in India, it is not likely that in this country complete enumeration will be entirely dispensed with, but rather that it will be used as a basis upon which sampling surveys may be developed.

Professor Mahalanobis has laid particular stress on the results of the Bihar survey shown in Table 1, and it is to this that I should like to confine my final remarks. He says that his system of sampling was very successful, in that in the case of rice and pulses a very small sample of about 0.025 per cent gave results within about 2 per cent of the final result obtained from the total sample. We are not told what was the result of the small sample in the case of sugarcane. But what strikes one most about Table 1 is perhaps the variations between the various estimates, particularly that between estimates 8 and 9, which suggest something more than sampling errors. Estimate 9 for rice gave 13,650,000 acres, a fall of 355,000 acres compared with estimate 8. It is noticeable that this is a larger difference than that between estimate 1 (based on a sample of one in 4,175) and estimate 9, although it resulted from a large increase in the sample (from 80.7 to 100.0 per cent, or 19.3 per cent of the whole sample). If my arithmetic is not wrong, it seems that the total acreage in or about January, if it had been estimated from this 19.3 per cent of the total sample, would be only about 12.2 million acres.

What is the explanation of such large discrepancies? We have not been given sufficient detail to decide this point. An apparent explanation would be that there are what Professor Mahalanobis referred to as physical fluctuations, which would presumably mean that the acreage under rice changes substantially from month to month, or even from day to day. But this can hardly be the explanation, for if it were, any similarity between the first estimate in October and the last in January could hardly be attributed to the merits of the sampling methods employed. The question would also arise of what was being estimated : was it the acreage under the crop in October, or in January, or the total cropped acreage for the season; and how are any of these to be estimated from a sample drawn over a lengthy period? As for the remaining crops, estimate 8 for pulses is 2,963,000 acres, as against 2,903,000 for estimate 9, and the sugarcane estimates show similar intriguing differences.

I have no doubt that there is some very simple explanation of all these divergencies, and that this would have been made quite clear if circumstances had permitted Professor Mahalanobis to submit his paper in the customary form for printing and distribution to Fellows at this meeting.

Time will permit only one further very brief point. Professor Mahalanobis said that at the last stage, at the request of the Government, an estimate of 312,700 acres was also given for the area under potatoes, only about 0.26 per cent of the total geographical area covered by the sample survey, and he added that this supplies a good idea of the wide range of the survey. It would therefore be interesting to have details of this estimate and to know the sampling errors involved, how they were calculated, and how they compare with those of the other estimates.

Dr. George concluded by expressing the deep pleasure it gave him to second the vote of thanks for such a valuable paper from so distinguished a visitor.

The vote of thanks was then put to the meeting, and carried unanimously.

Dr. M. S. Bartlett, after adding his personal thanks and his welcome to Professor Mahalanobis, said that there was one incidental point raised by the last speaker, Dr. George, on which he also was not very clear, and that was the precise meaning of "physical fluctuations" as distinct from "sampling variations."

When considering the paper generally, what he felt was perhaps not so much the novelty of the methods used, but the value of the integration or synthesis of methods used to meet a variety of demand. It was rare to find one department which used these methods for such varied activities as agricultural crops and the investigation of the cost of living and family budgets. Possibly in this country, especially on the social and economic side, those concerned with similar investigations would be studying further the methods which Professor Mahalanobis had described. In particular, there was a great deal to be said for this study of variability, a point which was well recognized in agricultural work, but which had been brought out that evening on the social and economic side. It not only enabled one to judge the accuracy of one's averages at the end, but it gave invaluable information on the material itself.

To take an example. Supposing they made an enquiry into the consumption of a particular food in connection with rationing. Such a study of variation, not only between the broadly classified groups of consumers, but within those classes, would not only give the accuracy of the mean level as shown by the sample, but it would give the variation between those classes. A knowledge of such variation was desirable in order to estimate the change in the mean level of consumption resulting from rationing and the amount of individual hardship which would be entailed if the people were rationed at definite mean levels.

There was one further theoretical point he wanted to mention. Although it was not covered by the paper that evening, it was mentioned in the paper in the *Philosophical Transactions* which had been referred to. He thought it had some relation too with the last Table which Professor Mahalanobis had produced, which he gathered had something to do with the question whether units were random and independent or not. In his paper in the *Philosophical Transactions* Professor Mahalanobis referred to what he called space correlation functions; this was theoretically an interesting concept related to the time correlation functions which occurred in time series. These had been encountered before; they had, for example, been used by Sir Geoffrey Taylor in the theory of turbulence. The property of stationarity in time series was the property of homogeneity in space correlation functions. Further questions that arose were: Were these functions isotropic?—that is, independent of the direction in which one was measuring the correlation—the answer in agriculture was probably "no"; and Did these functions ever show periodicity?

The sort of question which one asked in agricultural design work was usually answered on the basis of a uniformity trial, from which was worked out semi-empirically what was the optimum design. The same answer (for constant variability) must follow from a knowledge of the correlation function. Such an alternative approach did not of course affect the final analysis of any data which was accumulated on a basis of the design finally decided; that followed the standard analysis of variance along the lines that had been described.

The following contributions were received in writing after the meeting:

MAJOR M. D. W. ELPHINSTONE: Professor Greenwood mentioned his own admiration for Professor Mahalanobis's mathematical skill; I also worship from afar. But even more do I admire the courage which has enabled Professor Mahalanobis to build up his organization in the face of all the discouragements and difficulties which he has had to face.

I have during the war been responsible for certain administrative statistics in G.H.Q., India, and I can imagine some of the difficulties overcome.

In the first place, there is apathy—an apathy difficult to imagine even in our own constant struggle against wilful ignorance at home. Even those who care for good work are apt to be worn down by the continuous output of energy required to get any new idea accepted. The strain is greater in a hot climate, and appears to affect white and brown-skinned races alike.

Then there is the temptation of intellectual dishonesty. Professor Mahalanobis gave examples throughout his paper: the temptation of the field worker to "fudge" his figures, the political pressure resulting in an inaccurate census, the hint of the virtual suppression of an unpalatable report. If Professor Mahalanobis is not absolutely honest in his own work, then he must have practised on us this evening one of the subtlest deceptions of his career. Heat engenders fixed opinions and unreasoned prejudices, and I know how easy it is to hunt for figures to prove a theory rather than to hunt in the figures for the truth.

Then there is the administrative skill and thoroughness with which his investigations are planned. His account of the size of his organization must have made some of us jealous, but he is right to emphasize the factors of time and cost even in his Bengal surveys. Bullock-carts are a slow way of moving about; his field workers must spend much of their time travelling. Then he let fall a mention of work done near Karachi—only some 1,500 miles away. Only by the most careful economy can so much work be done in such conditions by so small a staff.

To give more point to my praise, I may perhaps add that having discarded one of my staff for "negative efficiency" I felt that my section was shaping well. We carried out a small survey on the output of work by coolies, and I designed most carefully all the working sheets and rules for checking. When the figures came up to me, it appeared that the discarded one's successor (specially picked as likely to do well in such work) had not troubled to see that decimal points were in the right place, being of the opinions that (1) I should not notice and (2) that even if I did it would not matter, as the mistakes would average out. Not all people are like that in any part of the world, but it is clear that Professor Mahalanobis has had enough of them to have had to devise techniques for protecting himself against them, and, moreover, has used by-products of these techniques to yield useful statistical information.

Anywhere in the world his achievements would be noteworthy, but as Professor Greenwood admires his work in Combinatory Analysis, so I, who have had some very small experience of the difficulties he has had to face, admire the skill with which administrative troubles unconceived in this country have been overcome.

MR. D. M. SEN : The speaker has emphasized, and Dr. Yates has reiterated, the difficulties of dealing with recording mistakes in sample surveys in different countries. Dr. Yates has introduced into the discussion the much-abused concept of the "intellectual competence" of the field investigators. I have been associated, for some time now, with large-scale sample surveys in Great Britain. Recording mistakes are not a regional phenomenon, not even in their scale. They are common wherever large-scale surveys are undertaken. It would be totally wrong to try to correlate recording mistakes with "the state of intellectual competence," even if "intellectual competence" could be measured and calibrated, or averaged and graded, for countries as a whole.

To counter recording mistakes, various checks are necessary. I may perhaps be allowed to give some instances from my own firm, which is thought to be one of the largest sample survey organizations in the world. Out of the ten departments which share, between them, the whole work of these sample surveys, beginning from the collection of the data to their presentation to clients, two of the largest departments are charged with the functions of "checking field data" and "insepcting processed data." The Tabulating Department, moreover, applies several checks on its own work (punching etc.), and on Checking Department's work, which has supposedly already checked the data recorded by field investigators. I remember a somewhat irritated fieldman asking the question, "who checks Checking Department's work?" As a matter of fact, the Tabulating Department itself checks Checking Department's work to an extent, as I have mentioned above, but there are other departments too, which are doing exactly the same thing at various stages of processing the data. The Tabulating Department, for instance, uses the "unit" check on the Hollerith to ensure that purchases plus previous stocks equal sales plus present stocks on each card. The cards may have been checked rightly from the original information sheets which were vetted by Checking Department previously; yet it is felt necessary to employ the "unit test." Similarly for "unit costing." The fieldman records purchases in units and the total value. The cost per unit is computed, which serves as a check on the accuracy of fieldmen's records, since this unit purchase price is compared with the unit purchase price of other shops for the same goods or similar goods, and also compared against the "unit cost" price as recorded a month or so previously in the previous period's survey. This cost per unit is used by Tabulating Department on the Hollerith. The units are multiplied by the "unit cost" to arrive at the total purchase price. These totals for different shops are added up on the Hollerith, and their grand total for each item or brand from the Hollerith serves as a check against the sterling total, which the Checking Department has got on their comptometers for the purchases of the same item or brand.

It is therefore clear that within departments themselves there are various checks. There are also many checks between departments. Data are also checked against previous averages, which serves as a rough check, if not a precise one.

In the field itself, besides fieldmen, there are field supervisors, who in their turn are controlled by the field manager. The field supervisors' work consists of no actual work except constant supervision over the fieldmen's thoroughness of collecting data.

Over and above these checks in the field, and during the time when the data are passed through the Tabulator and processed for presentation in the form in which they are wanted by our clients, there is another department—"Inspection Department"—whose job it is to check the purchase, sales, stocks, price and other tables for internal correspondence and plausibility, having regard to known data about market conditions, etc. We often discover at the very last stages mistakes which should have been picked up weeks before. That reflects neither on the field investigators' "intellectual competence" nor on the "intellectual competence" of the checking and supervisory staff within the office itself. These mistakes, we find, are inevitable, and hence they are allowed for. Extra staff are employed to see that they do not go through. There are, of course, cases when deliberate dishonesty is noticeable. I recall an instance where the field investigator, who was supposed to have gone to certain shops and recorded their stocks of certain branded goods, never in fact went there. He filled in the field information sheets at his seaside hotel. In certain cases, where the fieldman detects inaccuracy in purchase data or his stock count, he is supposed to inform the office about it. We in the office know that a percentage of shop data will, over a period, be marked as inaccurate by the fieldmen, and if any one fieldman's audits are extraordinarily accurate, from that very fact alone, we sometimes suspect that the audits are not being conducted as thoroughly as they might have been. This particular fieldman went so far as to advise us from time to time that such and such shops should be thrown out of the sample for that period owing to inaccurate records, etc., and thus our confidence in his work was kept up. He was at last found out by certain discrepancies in his stock position which did not tally with our knowledge of the then existing zoning scheme for the distribution of certain products. There are thus instances of deliberate dishonesty which one can quote.

It is also very easy to question the "intellectual competence" of field investigators. In our work, for instance, present stocks are subtracted from past stocks by fieldmen themselves, so that if present stocks are higher than past stocks, the difference is shown up as a negative quantity. Obviously, this increase in stock, or this negative quantity, must be covered by an equivalent, or higher, purchase figure. It is physically impossible for stocks to increase without equivalent purchases, unless the previous stock-taking is erroneous. Now, one might say that all fieldmen should be intelligent enough to see that these minus quantities are covered by equivalent purchases. In fact, in a very large number of cases honest and very competent fieldmen fail from time to time to account for increase in stocks. Perhaps they would be called "physical fluctuations," but anyhow physical fluctuations are very difficult to explain in this context.

One word about the size of samples. The speaker has indicated that he often takes samples with a sampling fraction of 1 : 4,175, and his largest samples have been of the order of 1 : 168. The samples with which we deal in our particular field of Food and Drug trades, and which we know, from experience over a period of time, yield satisfactory results, are sometimes as large as 1 : 50 and even 1 : 25, but sometimes as small as 1 : 250. I should have thought the size of samples would be related to the variability of the data, among other factors, and consequently the sampling fraction is perhaps not in itself as significant as might be thought.

PROFESSOR MAHALANOBIS made a brief reply at the Meeting, and amplified it later, as follows :—

I am extremely gratified by the kind way in which my talk was received. Unfortunately, the paper could not be circulated nor presented before the meeting in a written form. This naturally left many gaps, to which references were made in the course of the discussion. Relevant material was in certain cases actually with me, but I had no time to discuss it at the meeting.

I greatly appreciate Dr. Yate's emphasis on the importance of close contact between mathematical statisticians and field workers and the need for intimate contact with numerical material. This is a point which I have always striven to emphasize in our Institute. My own view has always been that no one should be considered to have qualified as a statistician without having gone through an apprenticeship as a computer.

I am particularly interested in Dr. Yates's remarks about bias in estimating crop yields. I entirely agree with him in thinking that such troubles can probably be eliminated if a permanent organization of trained workers can be built up. Our problem has been to find out what can be done until such trained workers become available. I may also mention in passing that the size of the cut has been the subject of sharp controversy between the Indian Statistical Institute and the Imperial Council of Agricultural Research of the Government of India. I have been pressing for several years for a scheme to study and compare the differences in technique under the joint supervision of the I.S.I. and I.C.A.R., as this is the only way in which agreed conclusions can be reached. The proposal for joint investigations was strongly supported by Professor R. A. Fisher in his memorandum submitted to the I.C.A.R. in March 1945. Unfortunately, we have not yet succeeded in persuading the I.C.A.R. to agree to this proposal. I spoke to a number of statisticians in the United States a few weeks ago and I hope to take the matter up there again shortly.

Dr. Oswald George as well as Dr. Bartlett have asked what I meant by physical fluctuations. I have already referred to this point in the text. Conceptually, sampling variations may be considered to have arisen from the fact that only a small portion or a fraction (and not the whole) of the universe is being surveyed. Recording mistakes arise from the unconscious bias, or sometimes conscious and deliberate negligence or dishonesty, of the human agency. Even when these two sources of error are eliminated there still remains a certain margin of fluctuations which I have called physical. For example, it would not serve any useful purpose to try to determine the weight of a living human being correct to, say, a milligram. When the administrator demands an estimate for a certain entity which he calls "the acreage under rice in Bengal in a particular year" or "the per capita consumption of cereals, etc.," there is a certain margin of fluctuation which is residual in the dual sense of being something independent of sampling variations and recording mistakes and also of being of no importance for the practical purposes in view. Operationally, the different types of error are mixed up in practice; but it is possible to some extent to separate them by appropriate analysis. I may add that what I have in mind is something like the permissible margin of fluctuations in the quality control sense, where production may be considered to be under control even when the measurements do not strictly conform to statistical randomness. Determination of the limits of physical fluctuations (in relation, of course, to the purpose in view) is difficult and requires a purely experimental approach. We realize its importance, but cannot claim to have made much progress in the actual determination.

Dr. George referred to my use of the phrase "unitary and zonal, configurational," etc. I had given a preliminary account in the paper "On large-scale sample surveys" in *Philosophical Transactions*, and I had included a more up-to-date description of the terminology and notation of sample surveys in my original draft, which I had with me at the meeting, but to which I could not refer for lack of time. This has been included in the present paper.

Dr. George has pointed out the need for small sample theory. I agree, of course, that it might sometimes become necessary to use it. We found, however, that too fine a stratification was often uneconomical in the sense that the additional cost was not commensurate with the gain in precision. In fact, in large-scale surveys, we usually found large sample theory to be adequate for practical purposes. But I agree with Dr. George that in appropriate cases small sample theory would no doubt prove to be of great value. Incidentally, I may mention that one of the important lines of work in our Institute has been the development of the exact or small sample theory of distribution for multivariate analysis by my colleagues R. C. Bose and S. N. Roy.

I am glad to have Dr. George's support about the importance of errors other than sampling error. I also agree, that errors arising through faulty working of machines cannot be neglected. I did not refer to this, simply because there was nothing peculiar in the Indian experience to make it worth mentioning.

Dr. George has referred to my remarks on the defects of complete enumeration. Certain observations which I had no time to make at the meeting will be found in the text of the paper. I may add one or two more remarks. Complete statistics are often obtained almost as a by-product of routine administration—for example, the number of criminal cases sent up for trial. Apart from mistakes in compilation or similar clerical errors I have no doubt that such statistics would be substantially accurate. Consider another example, which is nearer to the subject of the present paper. In temporarily settled provinces of India, crop records are compiled by a permanent, and often hereditary, staff of village recorders (called *patwaries*) primarily for revenue purposes, as the assessment is fixed on the basis of crop production. In such provinces, crop-acreage figures are claimed to be sufficiently accurate for all practical pruposes. I have no personal experience of the quality of the work done by *patwaries*, but it may very well be true that crop records prepared by them are substantially accurate. This is quite different from what happens in a permanently settled province such as Bengal or Bihar. I had in mind the so-called complete enumeration made by an *ad hoc* staff whose number is large (of the order of fifteen or twenty thousand) and who are scattered widely over a province like Bengal. All available evidence points to the fact that work done under such conditions is unreliable. What I had wanted to emphasize was that the fact that complete paper records had been made ready was no guarantee that such records would be reliable. I may perhaps quote one concrete example.

In 1939 the Government of Bengal decided to prepare a complete record, plot by plot, of the land sown with jute. After these records were prepared the Government arranged to have certain portions checked by permanent Government officers. The primary records, when checked, were found so unreliable that the Bengal Government ordered all the records to be destroyed.

I agree with Dr. George who thinks that suitable systems of sample checks and controls can be used with complete enumeration with great advantage. Some of the checks described in this paper may be quite suitably taken over.

Dr. George has referred to the results of the Bihar crop estimates. He is quite right in pointing out that the area under any given crop is not determinate unless the date is specified. I have discussed this point elsewhere.[1] In a province like Bengal no machinery exists for obtaining the crop acreage on a specified date. The estimate obtained by a sample survey spread over a lengthy period must necessarily be somewhat indeterminate. In Bengal, by the alternative method of obtaining an estimate of complete enumeration, the result would be equally indeterminate, for the simple reason that the complete enumeration would have to be made by an *ad hoc* staff whose work would be spread over more or less the same period as that of a sample survey.

A little consideration would, however, show that although a specified date makes the crop acreage determinate it renders the estimate of total production necessarily somewhat vague. Whatever may be the crop acreage on a specified date it is likely that the harvested area would be somewhat different. As already mentioned, such differences are likely to be quite appreciable in India, especially in the case of rain-fed crops. In this situation we thought it advisable to develop a method for ascertaining the total production by a direct estimation of production per grid. This seems to be as far as one can go in existing conditions in Bengal.

I may mention here one particular procedure with which we have been experimenting for some time. At the time of harvesting, the yield per acre is determined by crop-cutting experiments in a number of grids. The total cropped area in each grid is estimated at the same time, and in this way the total yield for each such grid is ascertained. This enables a direct estimate to be made of the total production in each cell, zone or district, or in the province as a whole. All lands which go out of cultivation would not in this method contribute anything to the estimate.

In showing the progressive estimates for the *ad hoc* crop survey my object was to show the kind of data supplied to administrators. The sampling at each stage was not individually or independently representative of the whole "universe" as spread over space and time. Suitable modifications in the theory of estimation to cover such cases have not been worked out. Different progressive estimates necessarily referred to different regions of the province at different times, and were not strictly comparable. Differences were due to sampling variations and partly to actual physical differences in crop acreage arising from the different geographical coverages at different periods of the survey. My real point was that, in spite of such intrinsic differences, the figures supplied to administrators were probably quite useful for administrative purposes.

---

[1] Presidential Address to the Section of Mathematics and Statistics, *Indian Science Congress*, Baroda 1942, p. 25.

Dr. George has enquired about the proportional sampling errors. These, naturally, were large when the total acreage or the proportion of land under a particular crop was small. I have discussed this point in greater detail for the Bengal crop surveys, for which material was available for a number of years. The Bihar crop survey was, unfortunately, terminated after one year and sampling errors had not been fully investigated. I hope, however, to give an account of the effective margin of error in crop surveys in a subsequent paper.

I am glad Dr. Bartlett has referred to what I called space correlation functions, which I mentioned in the paper in *Phil. Trans.*, although I had no time to speak on this topic at the meeting. Since that paper was written certain other concepts have been developed; and, as already mentioned, some theoretical work has been done by R. C. Bose. The subject appears to be of considerable interest; but, unfortunately, owing to heavy pressure of other work, it has not yet been possible to tackle it on fundamental lines.

It may be useful, however, if I add a few brief remarks. Consider a field of, say, square or rectangular shape consisting of a very large number of rows and columns of elementary cells or quads. Along any single row (or column), the problem of space correlation would be one-dimensional, and identical with that of serial correlation in time series. The fact that space distributions are intrinsically two-dimensional, however, makes it necessary to have more generalized concepts. The approach at present adopted in the Statistical Laboratory is as follows.

Consider a single row or a single column consisting of a very large number of cells. (The boundary difficulty can be avoided by making the row or column endless by joining the two terminal cells.) Imagine any particular row or column divided into a number of segments, each segment itself still consisting of a very large number of individual cells or quads. The correlogram or serial correlation for different gaps can be then determined for each segment. The form of the correlogram or serial correlation for different gaps can then be determined for each segment. The form of the correlogram or the serial correlation function may then differ from segment to segment or may happen to be statistically the same for different segments. The serial correlation is usually called stationary if it happens to be statistically the same for different segments in whatever way the segments are formed (subject, of course, to the provision that each segment continues to have a sufficiently large number of quads to enable the serial correlation to be determined with the required precision).

Now consider the adjoining (or some other) row. The correlation function or serial correlation in this row may again be stationary, but may differ from the stationary serial correlation in the preceding row. Thus, although the serial correlation in each row is stationary, they may be all different. In a simplified case, the serial correlation along each row may be stationary and may be the same for all rows. Such a field may be called "uniform along rows."

Now consider the columns. The serial correlation along each or some of the rows differ from segment to segment. If they happen to be the same over all segments in any particular column, then the serial correlation is stationary over that column. In the same way, the serial correlations may be stationary for each column and finally may be the same for all columns, in which case the field would be "uniform along columns."

It is possible, however, to consider the serial correlation not merely along rows or columns but in different directions, making different angles with the direction of, say, rows. If the field is divided up into two-dimensional segments (each segment still consisting of a sufficiently large number of elements or quads) it is possible to determine the serial correlation in different directions. In this way it is possible to construct what may be called the surface of the serial or space correlation in such a way that a section of the surface in any particular direction supplies the corresponding curve of the serial correlation. If such surfaces of the correlation function happen to be the same for all segments, then the field may be called homogeneous. It might not, however, be isotropic. In fact, in each segment the field would be isotropic only if the correlation function is identical in all directions or, in other words, the surface of the correlation function is a surface of revolution.

The interesting fact then emerges that the surface may be isotropic for each segment, but the actual surface may be different in different segments, in which case the field would not be homogeneous, but would be isotropic. It is obvious that the field may be homogeneous, but not isotropic. Finally, when the field is both homogeneous and isotropic it may be called uniform.

Instead of working with the correlation function it is also possible to work with the variance function in the same way. As the two functions are mathematically connected the results based on them would be equivalent. For practical purposes the variance function would often be more convenient, as it usually enables the sampling errors to be calculated more directly.

I may add that the study of the surface of the correlation or variance function is of interest in connection with what I have called mapping problems, in which it is desired to ascertain the actual distribution of the variate under study over different elements or regions of the field under survey.

In the paper in *Phil. Trans.* I explained that fields may be of non-periodic, quasi-periodic or periodic types. I also explained the use of the patch number in characterizing the different types of fields. From the point of view of patch numbers, fields may be broadly considered as belonging to three types, namely, (*a*) fields in which the patch number is very small, corresponding to what I called "patterned" fields; (*b*) fields with patch numbers lying within an assigned range on either side of the model or expectation value, which I called fields of the random type, and finally; (*c*) fields in which the patch number was larger than the limit assigned for random fields, which I called the crystalline type and which would correspond to quasi-periodic and periodic types from the point of view of correlation function.

It is easy to see that previous knowledge of the type of the field can help in improving the efficiency of the sampling design, especially in the case of mapping surveys. For example, if the field is of a random type then a purely random sampling would be appropriate. If the field is however of quasi-periodic, or periodic, or what I have called of the crystalline type, then cluster or configurational sampling must prove quite useful. On the other hand, if the field is known to be of the patterned type then it should be possible to improve the efficiency by suitable zoning and stratification.

I am particularly grateful to Major Elphinstone for his kind remarks. I entirely agree with him in thinking that the apathy of administrators and the peculiar difficulties in which statistical work has to be carried out in India has to be experienced in order to be properly appreciated. The average administrator in India expects the scientific or statistical technician to supply evidence or proof in favour of what the administrator thinks to be right, rather than to give independent advice on objective grounds. Intellectual dishonesty, to which Major Elphinstone has referred, would in such circumstances be an actual advantage in securing promotion in official posts. This is why I have never favoured the idea of the Statistical Institute being run as a Government department or under predominating Government control. I think it is of the utmost importance that scientific organizations like the Statistical Institute should maintain an independent position, so as to be able to offer impartial technical service and advice on an objective basis. Mr. D. M. Sen has referred to the size of samples. I am afraid I had not made the position quite clear in the paper. It is not correct to say that we have used a sampling fraction of 1 : 4175. This was the fraction of the sample (in Table 1) on which we supplied a purely provisional first estimate. In the sample survey of crop acreage we are at present using a sampling fraction of the order of about 1 : 150 or 1 : 200. In estimating the yield of crops per acre the sampling fraction is necessarily far smaller and of the order of 1 : 6,000,000. On the other hand, in small-scale work one may go upto such large fractions as 1: 20. Mr. Sen is entirely right in thinking that the size of samples would depend on the variability of the material. Naturally, it also depends on the precision it is desired to attain in the final results, and on permissible budget limits.

NATIONAL REGISTER OF SCIENTIFIC AND TECHNICAL PERSONNEL (GOVERNMENT OF INDIA)

Government of India is compiling a National Register of all qualified Scientific and Technical Personnel of the country. The National Register contains comprehensive data regarding Indian Scientists, Engineers, Technologists and Medical specialists at home and abroad; supplies information about registrants to the employers on request; and provides technical man power data for planning.

Minimum qualifications for registration are :

(1) Master's Degree in a Science Subject (including Mathematics and Geo-Sciences).

(2) Specialisation in Medicine.

(3) Bachelor's Degree in Agriculture, Veterinary Sciences, etc.

(4) Diploma in Engineering or Technology.

Registration is free and Registration Card 'G' may be obtained from the National Register Unit, CSIR, Old Mill Road, New Delhi. Registration card may also be obtained in person from Public Service Commissions, Universities, Employment Exchanges, National Laboratories.

BIBLIOGRAPHY OF NONPARAMETRIC STATISTICS AND RELATED TOPICS

A revision is being made of "Bibliography of Nonparametric Statistics and Related Topics," published in *Journal of the American Statistical Association* (1953), **48**, 844–906.

Material through 1959 is to be included with more emphasis, it is hoped, on applications than previously. References (particularly to the non-English literature), reprints, and technical reports on the theory or applications of non-parametric statistics would be greatly appreciated. Also, corrections and additions to the original bibliography are desired. Materials may be sent to I. Richard Savage, Statistics Department, University of Minnesota, Minneapolis, Minnesota, U.S.A.

BUREAU OF EDNL. & PSYL. RESEARCH